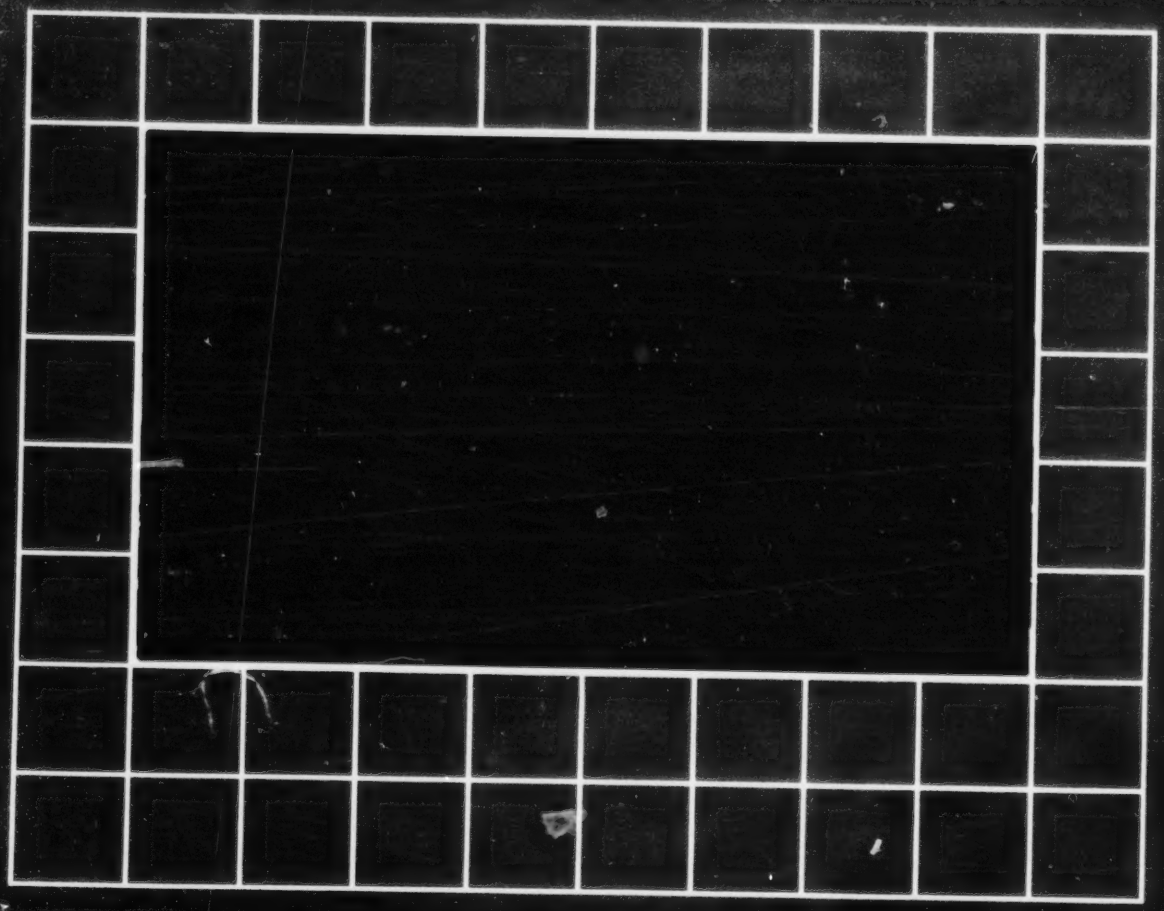
Employment and Immigration Canada

Emploi et Immigration Canada

COPS-Consultant Report



Canada

A REPORT ON A MODEL FOR MAKING PROJECTIONS OF INTERPROVINCIAL MIGRATION FLOWS IN AGGREGATE AND BY OCCUPATION

by

Bill Ahamad

and

Robin Rowley

Ahamad Consultants Inc.
17 Chesswood Court,
Nepean, Ontario.
K2E 7E3

24 January 1986

WH-6-106E

This report was prepared by Ahamad Consultants Inc. for the Labour Market Outlook and Structural Analysis Branch. The opinions expressed are those of the author and do not necessarily reflect those of the Branch or of the Government of Canada.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

I INTRODUCTION

I.1 Background and Objectives
I.2 Data

II AGGREGATE MIGRATION FLOWS

II.1 Introduction
II.2 Models
II.3 Factors Affecting Migration Flows
II.4 Regression Results

III OCCUPATIONAL DISTRIBUTION OF MIGRATION FLOWS

III.1 Introduction
III.2 Results

IV CONCLUSIONS

REFERENCES

APPENDIX

Some technical considerations for further analysis

TABLES INCLUDED IN THIS REPORT

TABLE II.1	PATTERNS OF RETURN MIGRATION FOR PERSONS WHO OUT-MIGRATED IN 1967-68, BY PROVINCE
TABLE II.2	PATTERNS OF OUT-MIGRATION OF PERSONS WHO IN-MIGRATED IN 1967-68, BY PROVINCE
TABLE II.3	PERCENTAGE OF OUT-MIGRANTS WHO RETURNED TO HOME PROVINCE ONE YEAR AFTERWARDS
TABLE II.4	PERCENTAGE OF IN-MIGRANTS WHO MOVED TO ANOTHER PROVINCE ONE YEAR AFTERWARDS
TABLE II.5	SELECTED MIGRATION EQUATIONS FOR 1983-84 - FEMALES AGED 18-24
TABLE II.6	SELECTED MIGRATION EQUATIONS FOR 1983-84 - FEMALES AFGED 25-44
TABLE II.7	SELECTED MIGRATION EQUATIONS FOR 1983-84 - FEMALES AGED 45-64
TABLE II.8	SELECTED MIGRATION EQUATIONS FOR 1983-84 - MALES AGED 18-24
TABLE II.9	SELECTED MIGRATION EQUATIONS FOR 1983-84 - MALES AGED 25-44
TABLE II.10	SELECTED MIGRATION EQUATIONS FOR 1983-84 - MALES AGED 45-64
TABLE II.11	EQUATION EXCLUDING LAGGED FACTORS FOR FEMALE MIGRATION IN SELECTED YEARS
TABLE II.12	EQUATION EXCLUDING LAGGED FACTORS FOR MALE MIGRATION IN SELECTED YEARS
TABLE II.13	EQUATION INCLUDING LAGGED FACTORS FOR FEMALE MIGRATION IN SELECTED YEARS
TABLE II.14	EQUATION INCLUDING LAGGED FACTORS FOR MALE MIGRATION IN SELECTED YEARS
TABLE II.15	EQUATION EXCLUDING EMPLOYMENT/LABOUR FORCE RATIOS FOR FEMALE MIGRATION IN SELECTED YEARS
TABLE II.16	EQUATION EXCLUDING EMPLOYMENT/LABOUR FORCE RATIOS FOR MALE MIGRATION IN SELECTED YEARS
TABLE III.1	OCCUPATIONAL DISTRIBUTION (%) OF TOTAL POPULATION FROM CENSUS AND FROM CPP
TABLE III.2	OCCUPATIONAL DISTRIBUTION (%) OF ALL MIGRANTS FROM CENSUS AND FROM CPP
TABLE III.3	OCCUPATIONAL DISTRIBUTION (%) OF MIGRANTS SINCE 1981 AND SINCE 1980 FROM CPP

EXECUTIVE SUMMARY

One of the objectives of the Canadian Occupational Projections System (C.O.P.S.) is to make projections of the demand and supply of manpower in different occupations and provinces. Interprovincial migration represents an important component of the supply of manpower, so that projections of migration flows need to be included in C.O.P.S. if the projections are to be useful for policy and planning.

This report describes a number of statistical models that have been developed for making projections of interprovincial migration flows in aggregate, and for disaggregating these aggregate flows by occupation.

Data on aggregate migration flows between provinces are available from a number of sources, but none is completely satisfactory for developing projection models. Data derived from taxation statistics by Statistics Canada appear to be the most useful at present since estimates are readily available for age and sex groups for each year since 1976-77. However, they suffer from the disadvantage that they are based on information for individuals who completed income tax returns, and the time period covered by the estimates is somewhat short for developing reliable models.

Our models for aggregate migration flows are based on the "gravity" model, which has been widely used in research studies on regional migration. The gravity model assumes that migration flows between two provinces (the "sending" and the "receiving" provinces), are related to their population sizes and to the distance between them. We modified this basic model to include measures of the "economic distance" between the two provinces as other explanatory factors: these include the difference between unemployment rates, the employment/labour force ratio and relative earnings.

We used the familiar regression framework for obtaining estimates of the parameters of different equations for interprovincial migration flows in given years. The results show that the most important factors affecting migration flows appear to be the population in the receiving and in the sending province, the distance between the two provinces, in-migration in the previous year (which we used to pick up some of the effects of return migration), and out-migration in the previous year (which we used to pick up the stability in migration flows). The estimates of the parameters for the economic factors were not statistically significant in some equations (which is in line with results of other studies on interprovincial migration), so that they should not be used in a projections model without further analysis. The models for the different years show a great deal of promise for projection purposes since the fit of the equations (as measured by the coefficient of determination, i.e. R-squared) is very good indeed. However, the estimates of the coefficients vary somewhat from one year to the next, and this suggests that more work is needed on the dynamic aspects of migration flows.

Data on the occupations of migrants are not available on an annual basis, but our analysis of data from the 1981 Census and from the Labour Force Survey Supplement in August 1982 suggests that the occupational distribution of migrants is reasonably stable over time. We therefore believe that reasonable projections by occupation can be derived by making assumptions about the occupational distribution of migrants from 1981 Census data. Some experimental analysis can be carried out to determine how sensitive the projections are to changes in these assumptions.

Our overall conclusion is that our findings provide a reasonable basis for developing projections of interprovincial migration flows. Our findings provide strong support for our basic methodology, despite the difficulties we have noted. The regression framework provides a short list of explanatory factors which, after further experimentation, may provide a satisfactory basis for making projections. At the same time, it seems important that further research should be undertaken to pick up the effects of economic factors and to improve the dynamic aspects of the models. Moreover, the data that are readily available for building projection models for interprovincial migration are not completely satisfactory, and improvements in the data base will probably generate significantly better models.

I INTRODUCTION

1.1 Background and Objectives

One of the objectives of the Canadian Occupational Projections System (C.O.P.S.) is to make projections of the demand and supply of manpower in different occupations and provinces. Interprovincial migration represents an important component of the supply of manpower so that projections of migration flows need to be included in C.O.P.S. if the projections are to be useful for policy and planning.

In order to determine the feasibility of building a model for making projections of interprovincial migration flows by occupation, we first carried out a background study to review the available data and existing methodologies for doing so, (see Ahamad, 1984). Our background study included a review of the literature to examine the results of recent empirical research on migration. This showed that some research studies have focussed on the importance of personal characteristics, such as age and sex, on interprovincial migration, while others have focussed on the effects of economic factors, such as income and employment. Both approaches have generated significant findings and we concluded that we should experiment with the inclusion of age, sex and economic factors in the model developed for C.O.P.S.

We also looked at the different sources of data that can be used for building a model. These include the 1981 Census, the Labour Force Survey Supplement carried out by Statistics Canada in August 1982, annual estimates derived by Statistics Canada from records of family allowance payments, annual estimates developed from taxation records by Statistics Canada and the Longitudinal File maintained by CEIC. None of these is completely satisfactory for making projections of interprovincial migration flows by occupation. Data on the occupations of migrants are only available from the Census and from the Labour Force Survey Supplement, but these provide only a static picture of migration.

We therefore concluded that our model for making projections of interprovincial migration flows by occupation would have to be built in two separate parts or sub-models. The first would consist of a model for projecting interprovincial migration flows in aggregate, while the second would consist of a method for disaggregating these flows by occupation.

We also concluded that our model for projecting interprovincial migration flows in aggregate should be based on the time-series estimates developed from taxation records by Statistics Canada since these are readily available. There are, at present, no annual estimates of the occupations of migrants so that the best that can be done for disaggregating these flows by occupation, is to base our projections on assumptions about the occupational distribution of migrants. We develop these assumptions by analysis of data from the 1981 Census and from the Labour Force Survey Supplement of August 1982.

To summarize, the objectives of this report are to describe the results of our research in developing a model for interprovincial migration for inclusion in C.O.P.S. More specifically, our objectives are to describe:

i) a model for making projections of interprovincial migration flows in aggregate using both economic factors, such as employment and income, and demographic variables, such as age and sex, as determining factors or independent variables;

ii) a model for disaggregating the projections derived in (i) by occupation using 1981 Census data and data from the Labour Force Survey Supplement of August 1982.

I.2 Data

Our data on aggregate migration flows were provided by the Administrative Data Development Division in Statistics Canada. These data are derived from personal income tax records and they are based on the province of residence reported on income tax returns. The data we obtained cover the annual migration flows between provinces for the period 1976-77 to 1983-84 by sex and by age-group (-24, 25-44, 45-64). Thus, the data available for our analysis consisted of 90 observations for males and females in each of the three age-groups, and for each of the eight years.

A number of limitations of these data should be borne in mind. Since our data on migration flows are based on taxation records, they cover only persons who filed a tax return. This group represents about 80 percent of the population (see Norris, 1983). Moreover, for the period up to 1980-81 the lowest age-group (-24) includes persons over 15 years of age; however, for the period since 1981-82, the lowest age-group includes only persons aged 18 years and over. It follows that the data are not strictly comparable over time.

Our data on the factors included in our models to project aggregate migration flows were obtained from two sources. Our estimates of unemployment rates, employment, labour force, and population by age, sex and province are derived from the Labour Force Survey, which is a monthly sample survey of about 56,000 households (see Statistics Canada catalogue 71-001). The samples in some of the age-sex-province groups (for example, for females aged -24 in Prince Edward Island), are so small that some of the data could not be obtained for them. We obtained our estimates for the relevant years directly from the Labour Force Survey Sub-division in Statistics Canada.

Annual estimates of average earnings are available from the Surveys of Consumer Finances (SCF) conducted by Statistics Canada. In some years, the SCF was carried out as a supplement to the Labour Force Survey, and the sample was large enough to generate earnings data of adequate reliability by age, sex and province, (see Statistics Canada, 1984). This procedure was used in developing the earnings estimates for 1982, 1981, 1979 and 1977 among the more recent years. However, in other years (e.g. 1980 and 1978) the SCF was based on a small independent sample, and only limited earnings data are published for these years. Most of our analysis was therefore carried out with data for 1982, 1981 and 1979. We obtained our earnings data for each year directly from the Household Surveys Division in Statistics Canada.

Data on the occupations of migrants are available from two sources: the 1981 Census, and the Current Population Profile (CPP) which was carried out as a supplement to the Labour Force Survey in August 1982. Census data on the occupations of male and female migrants have been published by Statistics Canada, but only limited data have been released from the CPP. However, we obtained the required data from special runs on the master file carried out by the Special Surveys Division in Statistics Canada.

The migration data available from the 1981 Census and from the CPP are somewhat different. In the Census, individuals were only asked to report their province of residence five years before, while in the CPP respondents were asked to report their province of residence on 1 June 1976 (i.e. about six years before the survey), as well as the dates on which they had moved since that time. The data we obtained from the CPP cover the occupational distributions of individuals who had changed occupation a) since 1976, b) since September 1981 and c) since September 1980. Because of the size of the sample used in the CPP, data were only obtained for the occupation major groups (two-digit code in the Standard Occupational Classification).

II AGGREGATE MIGRATION FLOWS

II.1 Introduction

The data on aggregate migration flows from taxation records can be used in two possible ways to develop a model for making projections. One possibility is to build a time-series model in which the migration flow in each year is treated as one observation. The advantage in this approach is that the resulting model will reflect the effects of changes in migration over time. The disadvantage is that since only eight observations are available, the best that can be done is to extrapolate migration rates over time. Thus, if we were to build a time-series model, we would not be able to include the effects of factors thought to affect migration, such as unemployment rates and incomes. This means that the model could not be used in experiments to determine the effects of changes in these factors on migration.

The other possibility is to build a cross-sectional model in which gross migration from one province to another is treated as one observation. The advantage of this approach is that we then have 90 observations in any given year for estimating an equation for making projections, so that we can estimate the effects of factors such as unemployment rates and incomes on migration flows. The disadvantage of this approach is that the effects of these factors may change over time so that the model for a particular year may not be completely satisfactory for making projections. However, since data are available for eight years, models can be built for different years to see if the estimated coefficients do change over time, and to determine if such changes in the coefficients have a marked effect on the projections.

We believe that the second approach is more in line with the general approach in C.O.P.S., which allows users to derive alternative projections using different assumptions about key factors. Thus, we based our models on cross-sectional data for a given year but we fitted equations for different years so that changes in the estimated coefficients over time can be examined.

We used linear regression analysis as our basic statistical method for developing equations for making these projections. We included the dependent variable in alternative forms in different equations to determine which would be most appropriate for developing projections for C.O.P.S. For example, in some equations we used the number of migrants between any two provinces, while in others we used the number of migrants standardized with respect to the population in the sending province (province of origin) and in the receiving province (province of destination).

Our review of recent migration research (Ahamad, 1984) suggested that migration varies considerably by age and sex. We therefore fitted separate equations for males and females in three age-groups: -24, 25-44, and 45-64. This means that we fitted six equations for each form of dependent variable in each year.

We included a number of factors or independent variables in different equations on a trial basis to determine which provide the best statistical explanation of the variation in migration flows. These factors include: employment/labour force ratios, the difference between unemployment rates, relative earnings and the distance between provinces (which is used as a proxy for moving costs). Since some migrants return to their province of origin shortly after migration, we included in-migrants in the previous year as one of our explanatory factors. We also included out-migrants in the previous year as another explanatory factor, since it seems reasonable to assume that some proportion of migration is relatively stable over time.

II.2 Models

We use the familiar "gravity" model as our basic framework for developing a model for projecting aggregate migration flows. This model gets its name from the fact that migration is assumed to be related directly to the size of the population in the sending province (province of origin) and in the receiving province (province of destination), and to be inversely related to the distance between them. Economic factors are also often included in the model to capture the effects of "push" and "pull" factors, (see for example, Greenwood, 1975).

In some applications of the gravity model, the migration flow from province "S" (for sending province) to province "R" (for receiving province), MSR, is first standardized with respect to the population of the sending and receiving provinces, and it is then linked through multiple linear regression to a number of explanatory factors, such as interprovincial differences in unemployment rates, in earnings and in the employment/labour force ratio. The difficulty with this approach is that the standardization procedure may take different forms. Some researchers therefore prefer to include the population in the sending and receiving provinces as explanatory factors, so that they appear on the right-hand side of the equation. We experimented with both possibilities in this study.

The gravity model is essentially multiplicative in its mathematical form, but we used a logarithmic transformation to make it linear and hence compatible with the familiar linear regression model. In exploring the usefulness of different specifications, we used three different forms for the dependent variable:

i) Log MSR
ii) Log MSR - Log POPS - Log POPR
iii) Log MSR - Log (POPS + POPR)

where POPS and POPR are the population in the sending and receiving provinces respectively.

We also used a number of factors or explanatory variables (which are described more fully in the next section) in one or more of our models:

i) UNR - UNS
ii) EMPS/LFS
iii) EMPR/LFR
iv) Log AVEYR - Log AVEYS
v) DISTSR

where UNS (UNR) is the unemployment rate in the sending (receiving) province, EMPS (EMPR) is employment in the sending (receiving) province, LFS (LFR) is the labour force in the sending (receiving) province, AVEYS (AVEYR) is average earnings in the sending (receiving) province, and DISTSR is an indicator of the distance between the sending and the receiving provinces.

We also explored the possibility of using a direct estimate of return migration as another factor in our equations. However, data obtained from the Administrative Data Development Division in Statistics Canada suggest that patterns of return migration vary considerably by province and by the number of years since initial migration. Thus it was difficult to develop a simple indicator of return migration, and we tried to capture the effects of return migration by including the number of in-migrants lagged by one year as another explanatory factor.

II.3 Factors Included in our Models

a) Population in receiving and sending provinces (POPR and POPS)

As noted in the preceding section, the gravity model assumes that migration flows depend directly on the population in the sending province and the population in the receiving province. This clearly makes sense in terms of the sending province: for example, if we were to assume that the probability of out-migration is fixed (other things being equal), then it follows that the number of out-migrants would be greater for a large province than for a small province.

It is less obvious that migration flows should also be related to the size of the population in the receiving province. However, we can think of the population in the receiving province as a proxy for the size of the labour market there. Employment opportunities and information about job vacancies will tend to be higher in large provinces than in small provinces, so that large provinces will tend to attract more migrants. Note that our data are specific for the six age-sex groups, which seems essential for these arguments to be valid. They are weakened by the potential coincidence of retirement and migration for some members of the oldest group.

b) Distance between provinces (DISTSR)

In the gravity model, distance is assumed to act as a major deterrent to migration, and several explanations have been advanced to justify its inclusion in behavioural models. These include the cost of information and of transportation, and the psychic costs of moving. Thus we expect that migration will tend to be higher for provinces that are close together, than for provinces that are far apart. It follows that the sign of the distance variable in our regression equations should be negative. The psychic or implicit costs of moving might reasonably be considered to be different for the different age-sex groups because of differences in family responsibilities and other personal pressures. This suggests that the coefficients for distance will vary across groups.

Some analysts have used the mileage between the capital cities of the various provinces as the measure of distance included in their model, but this assumes that migration is linearly related to distance. We have used instead a three-point scale to indicate distance: 1 for migration across one provincial boundary or within the Maritime Provinces, 2 for migration across two provincial boundaries or between Newfoundland and any of the Maritime

Provinces, and 3 for all other migration. Our assumption is that although distance plays an important part in migration decisions, its effects become fixed after a certain point.

c) Relative earnings (AVER/AVES)

It seems reasonable to assume that migration decisions will be highly influenced by the income that individuals expect to earn when they change province. In turn, the earnings that individuals expect when they change province, depend on factors such as their level of education and the relative demand for their skills in both provinces. Highly educated persons and those with skills that are in short supply, are generally more likely to migrate than those with limited education or skills, since there is often a good market for their services outside the local area.

Time-series data on the education and skills of migrants are not currently available so that we could not explore the full effects of earnings differences between provinces in developing our model for projecting migration flows. However, we included relative average earnings in the receiving and sending provinces as one of our explanatory variables, in an attempt to capture some of the effects of earnings differences.

d) Difference in unemployment rates (UNR - UNS)

Many empirical studies on internal migration in Canada and in the United States have included unemployment rates as one of the explanatory factors. It seems reasonable to assume that areas with high unemployment rates will experience both higher out-migration and lower in-migration than areas with low unemployment rates. Somewhat surprisingly, however, the effects of unemployment rates have not usually turned out as expected, (see for example, Vanderkamp, 1973).

Different explanations have been advanced for the failure of empirical analysis to pick up the effects of unemployment rates, (see Greenwood, 1975). Some economists argue that unemployment is both a cause and an effect of migration, so that the estimated coefficients in regression equations will be biased. Others suggest that unemployment has, until recently, affected a very small part of the population, so that unemployment rates may not be an important factor for the bulk of the population.

Despite these difficulties, we decided to include the difference between the unemployment rates in the receiving and sending provinces as one of the explanatory factors in our regression equations.

e) Employment/labour force ratio in receiving and sending provinces (EMPR/LFR and EMPS/LFS)

Following the work of Green (1977), we also included the employment/labour force ratio in the receiving province and in the sending province as additional factors in our regression equations. These factors are related to the unemployment rate in each province since the latter is (1 - employment/labour force) 100. Thus the difficulties discussed above for unemployment rates apply equally to the employment/labour force ratios.

f) In-migrants in previous year (LMIGRSL)

Recent empirical research suggests that previous migration experience may be an important determinant of migration flows, (see, for example, Grant and Vanderkamp, 1976). Migrants are sometimes disappointed after their move, and this may encourage them to return to their home province ("return" migration), or to move to a third province ("onward" migration).

Data obtained from the Administrative Data Development Division in Statistics Canada suggest that return and onward migration are indeed important components of migration flows. The data in Table II.1 show the percentage of out-migrants in 1967-68 who returned to their home province in each of the nine subsequent years. For example, of those who left Newfoundland in 1967-68, 21 percent returned in 1968-69, 8 percent returned in 1969-70, 6 percent returned in 1970-71, and so on. These data therefore suggest that most of the migrants who chose to return to their home province, did so in the first year or two after migration. However, the percentage varies considerably by province: for example, the percentage who returned one year after initial migration varies from a high of 21 percent for Newfoundland to a low of 8 percent for Saskatchewnan.

The figures for those who moved again after in-migrating in 1967-68 (Table II.2) are somewhat higher than those in Table II.1, but the pattern of variation is similar: of those who moved into Newfoundland in 1967-68, 31 percent left in 1968-69, 16 percent left in 1969-70, and 5 percent left in 1970-71. The percentage who left in the first year after migration varies from a high of 31 percent for Newfoundland to a low of 13 percent for British Columbia.

It is also interesting to look at the changes in these percentages over time. In Table II.3 we show the percentage of out-migrants in different years who returned to their home province in the year immediately after migration: these percentages are shown for each year for the period 1967-68 to 1975-76, which are the only years for which the data are readily available. It is clear from the entries in the table, that the rates of return migration vary considerably from one year to the next for some provinces. For example, the percentage for Newfoundland varies from a high of 21 percent in 1967-68 to a low of 11 percent in 1973-74.

Data on the percentage of in-migrants who moved to another province in the year immediately after migration are shown in Table II.4 for each year in the same period. Here again we note that the percentage varies substantially over the period for some provinces. For Newfoundland, for example, it varies from a high of 31 percent in 1967-68 to a low of 19 percent in 1973-74.

Two main conclusions follow from our examination of these data. First, it is clear that return and onward migration are so large that some attempt should be made to include their effects in our projections model. Second, rates of return migration and of onward migration vary substantially by province and by year of migration, so that it is fairly complex to include them in a model. The data available to us were somewhat out-of-date, and a thorough analysis using more recent data is necessary before these factors can be included in a realistic way in a model.

As an approximation, we included the number of in-migrants in the previous year in each province as one of the explanatory factors in our regression equations. One disadvantage of this approach is that this factor will pick up the effects of return migration in the year immediately after initial migration but not in subsequent years. Another disadvantage is that this approach assumes that the effects are the same for each province, which as demonstrated by the above data, is clearly not the case. However, we believe that using this factor will provide a useful approximation until more research on return and onward migration can be undertaken.

g) Out-migrants in previous year (LMIGSRL)

As also noted above, the gravity model assumes that migration flows are determined by the population in the receiving province, the population in the sending province and the distance between the two. Since distance is fixed and the annual change in the population in the sending and the receiving provinces is relatively small, the model implies that migration flows will tend to be relatively stable from one year to the next.

In the context of the modified gravity model we are using here, we can think of migration as being made up of two separate components. One component depends on population and distance and so it is relatively stable; the other component depends on economic factors and it changes in response to changes in economic conditions in different provinces.

Some analysts have used a different argument to support their view that current migration flows are highly dependent on previous migration flows. Migrants are likely to transmit information to relatives and friends at home, and this will encourage greater migration to those provinces to which previous migration has been high. Moreover, individuals contemplating migration are more likely to move to a location in which they already have relatives or friends, since they may provide food, shelter and social contacts until the new migrant finds a job. Such individual situations introduce a heterogeneity which cannot be included in an aggregate model of the type being developed here.

The importance of previous migration in determining current migration has been clearly demonstrated in a number of empirical studies, (see Greenwood, 1975). The inclusion of previous migration in the model implies that the factors identified above have not only a direct effect on current migration, but also an indirect effect through previous migration. It follows that the inclusion of previous migration in our regression equations may have a substantial effect on the estimates of all of the other factors. In a later section, we indicate the need to find a scaling factor to permit comparisons of estimated regression coefficients when previous migration is included in the model.

TABLE II.1: PATTERNS OF RETURN MIGRATION FOR PERSONS WHO OUT-MIGRATED IN 1967-68, BY PROVINCE

Province	Percentage of Out-Migrants in 1967-68 Who Returned in:								
	1968-69	1969-70	1970-71	1971-72	1972-73	1973-74	1974-75	1975-76	1976-77
Newfoundland	20.9	8.1	5.9	3.7	1.8	2.9	1.8	2.9	1.8
Prince Edward Island	13.0	5.8	2.9	1.4	5.8	2.9	1.4	0.0	2.9
Nova Scotia	16.2	7.0	6.5	2.9	2.7	2.2	1.4	1.2	0.5
New Brunswick	11.7	9.2	4.5	3.5	4.2	1.7	1.7	2.0	1.2
Quebec	11.6	5.2	3.2	2.6	2.5	1.8	1.2	0.3	0.6
Ontario	14.9	6.8	5.9	2.6	2.5	1.5	1.0	1.0	1.1
Manitoba	10.0	3.9	3.1	2.9	2.4	1.1	1.0	1.1	0.4
Saskatchewan	8.3	4.2	2.6	2.8	2.2	2.0	2.2	1.4	0.9
Alberta	12.6	7.0	3.8	3.4	1.6	2.2	1.9	1.3	1.4
British Columbia	15.4	7.3	4.9	4.4	2.0	2.3	1.4	1.1	1.1
Canada	13.0	6.2	4.3	3.1	2.3	1.8	1.4	1.1	1.0

Source: Data obtained from Administrative Data Development Division, Statistics Canada.

TABLE II.2: PATTERNS OF OUT-MIGRATION OF PERSONS WHO IN-MIGRATED IN 1967-68, BY PROVINCE

Province	Percentage of In-Migrants in 1967-68 Who Out-Migrated in:								
	1968-69	1969-70	1970-71	1971-72	1972-73	1973-74	1974-75	1975-76	1976-77
Newfoundland	30.6	15.8	5.1	4.6	3.6	3.6	1.5	0.5	2.0
Prince Edward Island	23.0	9.8	14.8	4.9	3.3	1.6	4.9	4.9	1.6
Nova Scotia	21.7	9.8	7.5	4.1	5.4	2.1	2.1	1.6	1.3
New Brunswick	20.5	7.3	6.2	5.1	4.8	1.5	2.2	1.5	0.4
Quebec	19.2	10.6	8.6	5.9	4.4	2.5	2.2	1.8	2.0
Ontario	17.7	7.4	5.6	4.5	3.0	2.4	2.3	1.7	1.4
Manitoba	21.4	10.3	7.6	5.3	6.3	4.2	2.0	0.8	1.7
Saskatchewan	28.2	11.8	8.7	5.6	5.0	3.3	2.3	2.7	1.9
Alberta	16.8	8.8	6.6	5.2	4.1	2.9	2.2	1.8	1.2
British Columbia	16.8	8.8	6.6	5.2	4.1	2.9	2.2	1.8	1.2
Canada	18.8	8.8	6.4	5.4	3.8	2.5	2.0	1.6	1.3

Source: Data obtained from Administrative Data Development Division, Statistics Canada.

TABLE II.3: PERCENTAGE OF OUT-MIGRANTS WHO RETURNED TO HOME PROVINCE ONE YEAR AFTERWARDS

Province	Year of Out-Migration								
	1967-68	1968-69	1969-70	1970-71	1971-72	1972-73	1973-74	1974-75	1975-76
Newfoundland	20.9	18.0	18.1	18.5	18.6	15.0	10.8	17.4	17.7
Prince Edward Island	13.0	11.9	9.7	10.3	13.6	14.3	12.0	18.4	16.2
Nova Scotia	16.2	15.9	8.7	14.1	13.7	12.7	13.5	16.9	16.4
New Brunswick	11.7	14.2	14.9	15.6	13.5	15.2	13.2	15.1	13.1
Quebec	11.6	10.2	12.5	12.4	11.9	10.4	12.2	11.6	10.7
Ontario	14.9	14.9	14.7	13.8	15.0	13.6	13.7	13.9	14.7
Manitoba	10.0	10.4	12.0	14.0	10.6	12.7	13.8	14.3	12.8
Saskatchewan	8.3	9.4	9.5	10.4	11.1	12.8	16.2	11.2	13.0
Alberta	12.6	11.2	16.1	11.9	13.3	15.0	16.0	15.8	12.7
British Columbia	15.1	13.7	15.7	18.2	14.8	14.5	13.6	14.1	15.6
Canada	13.0	12.4	13.4	13.6	13.5	13.6	13.7	14.0	13.8

Source: Data obtained from Administrative Data Development Division, Statistics Canada.

TABLE II.4: PERCENTAGE OF IN-MIGRANTS WHO MOVED TO ANOTHER PROVINCE ONE YEAR AFTERWARDS

Province	Year of In-migration								
	1967-68	1968-69	1969-70	1970-71	1971-72	1972-73	1973-74	1974-75	1975-76
Newfoundland	30.6	23.7	23.0	22.4	25.8	32.7	19.4	26.7	27.4
Prince Edward Island	23.0	26.6	16.0	13.1	20.8	23.3	19.5	16.1	17.5
Nova Scotia	21.7	18.5	18.7	15.3	19.2	14.7	18.9	19.4	12.8
New Brunswick	20.5	20.2	17.9	18.5	19.5	18.4	19.9	20.8	21.6
Quebec	19.2	19.0	17.6	16.3	19.3	16.7	16.7	17.8	19.6
Ontario	17.7	15.1	17.1	17.1	17.7	18.6	17.8	17.9	17.7
Manitoba	21.4	19.7	23.6	20.7	20.7	20.6	21.8	20.1	27.5
Saskatchewan	28.2	24.4	27.3	22.8	23.0	19.9	19.2	21.3	21.4
Alberta	16.8	16.3	16.3	22.8	17.3	18.2	17.9	17.5	19.3
British Columbia	13.3	15.1	14.5	15.2	14.6	18.1	18.5	18.0	17.7
Canada	18.8	17.4	18.1	18.4	18.4	19.0	18.7	18.8	19.6

Source: Data obtained from Administrative Data Development Division, Statistics Canada.

II.4 Regression results

a) General discussion

The exploration of imprecise models for making projections of interprovincial migration flows is affected by a number of difficulties which cannot be completely resolved. These difficulties stem from: (i) reliance on aggregate migration data, which ignore the heterogeneity of individual experiences; (ii) the need to use measurable indicators which are assumed to be related to the motivation of individuals in choosing to migrate; (iii) the choice of functional relationships linking these indicators and migration flows; (iv) the use of annual data with no precise knowledge of the best interval for modelling migration flows; and (v) the discrepancies between the theoretical model used and its application in a real political and economic context.

These difficulties may have a considerable effect on the estimated coefficients of a regression model and they are important if the purpose of the model is either to estimate or to confirm the effects of the explanatory factors. However, they are much less important if the purpose of the model is to make projections. Experience with economic forecasting models suggests that models based on a combination of broad causal patterns and a simple model specification may provide better projections than models based on detailed theoretical considerations. For example, projections of economic variables based on simple time-series models have often proved to be better than those generated by large-scale econometric models.

This does not mean that we should ignore theoretical considerations and statistical tests in developing our model for making projections of interprovincial migration flows. Theoretical and empirical research suggest that demographic and economic factors are important determinants of migration and we therefore included them in our equations. But the confirmation of their influence was not our primary concern and we experimented with different equations to help us choose at least one which would provide reasonable projections of aggregate migration flows.

As noted above, we fitted a number of regression equations (over 2,200 in total) for migration between provinces for males and females in three age-groups and for six years. Our initial results encouraged us to make specific choices with respect to standardization for population and for the treatment of earnings differences between provinces. We then concentrated on fitting equations for the years 1983-84, 1982-83 and 1980-81 (i.e., the most recent years for which more reliable data are available).

Gravity-type regression models fitted by other researchers use a variety of methods for standardizing for different population sizes in different regions. As discussed above, (Section II.2), one choice is between: (i) adjusting the dependent variable by division or subtraction using a function of population size in the sending and receiving provinces; and (ii) including measures of population size as explanatory factors in the regression equation. Alternative (i) is more restrictive than alternative (ii) in that it contains prescribed values for the parameters and is therefore a special case of the latter. We explored both of these possibilities, as well as their reformulation using the employed population instead of the total population.

Several conclusions emerged from our initial experiments with different equations. First, the use of the sum of the population in the sending and in the receiving provinces is not as satisfactory as the inclusion of each as a separate explanatory factor in the regression equations. Second, when the lagged migration flows are excluded from the equations, the estimated coefficient for the two populations are different. Third, these estimated coefficients are different from unity so that modification of the dependent variable (instead of the free-parameter approach we used) does not appear to be appropriate. Fourth, it does not seem to matter whether we use total population or employed population for standardization. Finally, the inclusion of lagged migration terms increases the difficulty of appraising the need for standardization but the above four conclusions remain the same.

Some of our regression results are shown in Tables II.5 to II.16. These tables are organized in two parts: the first six tables show equations for the basic six-fold classification of migrants according to age and sex for 1983-84, while the second six are designed to focus on the intertemporal stability of the coefficients for three specific equations in each of three years.

One of the three specifications shown in the second set of tables includes the distance between provinces and for each of the six age-sex specific groups, the population and employment/labour force ratio in the sending and receiving provinces, relative earnings and the difference in unemployment rates between the two provinces. The other two specifications shown in these tables include interprovincial migration flows in the previous year for out-migrants as well as for in-migrants. As noted above, these factors are designed to pick up the effects of the stability of migration flows and return migration, respectively. They therefore introduce a dynamic element which is absent from the specification shown in Tables II.11 and II.12.

The coefficients of multiple determination (R-squared), which indicate the goodness-of-fit of the equations, are surprisingly high in view of the considerable range in the migration flows, (in 1983-84, for example, the flows ranged from a high of 6,227 for males aged 25-44 to a low of 3 for females aged 18-24). The lowest value of R-squared shown is 0.68, and the value increases considerably when the lagged migration flows are introduced into the equation. These high values for R-squared are very encouraging in terms of potential usefulness of the equations in making projections.

b) Results for 1983-84

The entries in Tables II.5 to II.10 show our regression results for 1983-84. The estimated parameters are shown for twelve different specifications for each of the six age-sex groups we analyzed. In the first four equations, migration is linked to the distance between the sending and receiving provinces, to their population size and employment/labour force ratios, to relative earnings and to differences in unemployment rates. With the exception of distance, the explanatory factors (or right-hand-side variables) are measured as of the preceding year, so that the equations have a recursive structure. This means that, given the values of the explanatory factors in a particular year, our equations can be used to project migration flows for the following year. Projections over a longer time-horizon need additional input since the values of the explanatory factors must also be projected or set at hypothetical levels before projections can be derived.

This recursive structure also exists in the other eight equations, which contain lagged variables representing the stability of migration flows (out-migration in the previous year) and return migration (in-migration in the previous year). Columns (5) to (8) inclusive show the estimated coefficients when out-migration in 1982-83 is included as an explanatory factor. In the next two columns, out-migration in 1982-83 is excluded but in-migration in 1982-83 is included. The final two columns show our estimated coefficients when both past out-migration and in-migration are included in the equations.

As pointed out in Section II.3, the inclusion of out-migration and in-migration in the previous year may have a considerable effect on the estimated coefficients in the equations. This is because the explanatory factors are then assumed to have a direct effect as well as an indirect effect (through previous migration) on migration flows. Thus the inclusion of lagged migration

flows may change the signs of the coefficients and hence the interpretation of the estimates. Unemployment rates and employment/labour force ratios are also inter-dependent so that there is also some interaction between these two factors. We followed the approach in many other research studies by using the difference between unemployment rates as an explanatory factor, but this factor ignores the size of unemployment rates in the sending and receiving provinces. The use of employment/labour force ratios assumes that the size of unemployment rates will also have some effect on migration flows, quite apart from the difference in rates between provinces.

Three other aspects of the data should be noted before we discuss the results. First, we fitted equations for three age-groups: -24, 25-44 and 45-64. These groups are consistent with the notion that there are distinct labour markets within the overall economy. Second, we used separate data for earnings and unemployment rates for males and females but not for the three age-groups. This is because the samples used for deriving these data for some age-sex groups are very smal; thus these data are subject to large sampling errors, and in some cases they are not published by Statistics Canada. Note that we used age-sex specific data for the other explanatory factors (population and employment/labour force ratios). Third, the use of a logarithmic transformation of migration flows creates difficulties if the size of the flow is very small. Thus we ignored very small migration flows (0, 1 or 2 persons) in estimating our equations. The effect of doing so is that we lost some observations for those aged 45-64. For example, for males aged 45-64, we excluded the migration from Prince Edward Island to Newfoundland. Despite such omissions, the data show considerable variation.

In the tables, we have used an asterisk (*) to indicate those coefficients for which the "t" value exceeds 2.0. As noted above, the use of statistical tests is not completely valid in the exploratory context we are using here, and the "t" values are used only as a guide in making a choice between equations. We also show the values of R-squared which indicate the goodness-of-fit of each equation.

The entries in the first four columns of each table (which exclude lagged migration) suggest that the coefficients of the population in the sending and receiving provinces are statistically significant. The estimated coefficients generally lie between 0.50 and 0.75, and there are slight differences between the two coefficients. Our overall impression (based on all of our estimates, including those not shown here) is that the coefficient for the population for the receiving province tends to be higher than that for the sending province.

The estimated coefficient for our measure of distance is also always significant, and the estimate falls within a fairly narrow range. Notice also that the estimate rises with age which is to be expected since the distance variable also reflects the cost of disrupting an established lifestyle. There are a few instances of sex disparities in the effects of distance. The coefficients for females are higher than those for males for the younger age-groups, but not for the oldest age-group.

The estimates for the economic factors are less satisfactory and none provides a firm basis for inference. This finding is in line with the results in many other studies on migration. As we noted above, the poor performance of economic explanatory factors may not matter for a projection model, but we believe that further experimental analysis using economic explanatory factors would be worthwhile.

The estimated coefficients for relative average earnings in the sending and receiving provinces tend to be smaller for males aged -24 and 45-64 than for the other groups. This is consistent with the views of some social scientists that average earnings are less important for these two groups than for others, and that family constraints are more important for females. These results provide some support for our use of separate equations for different age-sex categories.

The signs of the coefficients for relative earnings may seem somewhat strange since they suggest that migration to provinces with higher average earnings will be lower than that to provinces with lower average earnings. However, Miller (1973) has argued that negative coefficients are to be expected since high income areas have a more mobile population, while low income areas have a less mobile population.

The estimated coefficients for the difference between unemployment rates are fairly small. As noted above, the failure of migration models to pick up the effects of differences in unemployment rates has been something of a puzzle to economists and many explanations have been suggested (see Greenwood, 1975). In the context of Canada, it may be the case that provincial unemployment rates should not be treated symmetrically since information on them may not be equally available to all potential migrants. Moreover, the differences between migrants and non-migrants (for example, in terms of education) may be so large that provincial rates do not provide a good indicator of the situation as seen by the potential migrant.

The results suggest that the effects of employment/labour force ratios are unstable since the estimated coefficients change considerably as the model specifications change. One possible explanation is that this instability arises because of the interaction of this factor with the difference in unemployment rates, but this can only be a partial explanation. The estimated coefficients for these factors in column (3) in each table (which excludes the difference in unemployment rates) are always positive, and the coefficient for the receiving province exceeds that for the sending province in all cases except for males aged 45-64. This is consistent with the view that the "pull" of employment exceeds the "push" of unemployment. It suggests that differences in employment growth are more conducive to migration than differences in unemployment rates.

Differences in the coefficients of the employment/labour force ratios in the sending and receiving provinces are to be expected if unemployment rates are also included in the equation since these factors are inter-related. If we were to ignore the fact that we used different data to measure both factors (the employment/labour force ratios were based on age-sex specific data, while umemployment rates were not), it is easy to see that an equation which includes both factors can be manipulated to express migration as a function of only one of them: the new coefficients would then be a simple linear function of the estimated coefficients. This "compounding" of the coefficients makes it difficult to compare the estimated coefficients across rows in the tables, especially for the three labour-market factors. Moreover, conventional "t" statistics become less relevant when compound coefficients are the focus of interest.

This compounding phenomenon becomes more important for the estimated coefficients containing lagged out-migration, (columns (5) to (8) inclusive). The introduction of lagged migration in the equation means that the coefficients of the various factors cannot be compared directly with those in the other equations. However, we can apply a scale factor which makes comparison possible: it can be shown mathematically that all we need to do is to divide the coefficients in the equation with lagged migration by (1 - the coefficient for the lagged variable). For example, the coefficients in equation (5) in Table II.5 can be compared with those in equations (1) to (4) by dividing the former by (1 - 0.893).

The application of this scaling procedure suggests that the lower estimated coefficients for our measure of distance are quite consistent with the estimates that do not include lags. However, it is apparent from the entries in the tables that inclusion of lagged out-migration has a more complex impact than this simple scaling procedure would suggest. It follows that the dynamic aspects of migration need to be given more attention.

The high estimated coefficients for lagged out-migration, suggest that there is a marked stability in migration flows. As might be expected, the highest coefficients occur for those aged 25-44, but the coefficients are consistently over 0.8 so that their range is limited. The corresponding values for R-squared are high and they suggest that the equations are promising for making projections. If the other coefficients are scaled as suggested above, then it becomes clear that inclusion of the lagged variable does not have a neutral effect on the other economic factors. For example, if we compare the coefficients for relative average earnings in equations (1) and (7) in Table II.5, the scaled coefficient in equation (7) becomes 0.740/(1 - .902) which equals 7.551. It is therefore considerably larger than the estimated coefficient of 1.480 in equation (1).

Several conclusions follow from our examination of the equations with lagged out-migration: (i) lagged out-migration has a pronounced direct impact on migration flows; (ii) its inclusion is not neutral with respect to the estimated coefficients for the economic factors; (iii) the estimates confirm that the population in the sending and receiving provinces have different effects; and (iv) the results suggest that economic factors have important long-term effects, but more research needs to be undertaken to identify them.

The final four columns in the tables include the number of in-migrants in the previous year, which is intended to pick up the effects of return migration: out-migrants in 1982-83 are excluded from the equations in columns (9) and (10), but both in-migrants and out-migrants in 1982-83 are included in the equations in the two final columns. It is evident that in-migrants and out-migrants in 1982-83 are highly correlated, so that when one of these factors is excluded from the equation, some of its effects are picked up by the other. The estimated coefficients of lagged in-migrants in columns (9) and (10) are simply so large that they cannot be interpreted as reflecting only the effects of return migration.

The equations in the two final columns in the tables seem to be the most useful for making projections. They confirm the importance of the stability of migration flows and the importance of return migration - especially when they are considered with all of the other equations we fitted but which are not reported here. Both out-migration and in-migration factors are significant in terms of conventional statistical tests, and the size of the estimated coefficients appears to be reasonable in terms of the interpretation of these factors. We conclude that both of these factors would seem to be important in an equation to project aggregate migration flows.

We also conclude that our measure of distance between provinces provides another factor which should be included in a projections equation. The estimated coefficients for this factor are consistently significant using familiar statistical tests, and they appear to be reasonably robust to changes in the treatment of economic factors.

The population and the employment/labour force ratio in the sending and in the receiving provinces, provincial differences in unemployment rates and relative earnings, present more of a problem. They interact in the sense that inclusion or exclusion of employment/labour force ratios and of relative earnings changes the size of the coefficients for population. These factors may also interact in a manner that we have not been able to explore. Green (1977) in Canada and Moore (1973) in the United States are representative of economists who argue that labour market conditions may be better measured by the ratio of employment to the working-age population. Our results suggest that further exploratory analysis along these lines may be worthwhile.

The estimates in columns (11) and (12) in each table confirm that there is a need to treat the population in the sending and receiving provinces as separate explanatory factors. The estimated coefficients for the population in the receiving province range from 0.07 to 0.14, and they are positive and often statistically significant. By contrast, the estimated coefficients for the population in the sending province are smaller, sometimes negative and not statistically significant. However, the latter are sensitive to the exclusion of relative earnings, and they increase markedly and are often statistically significant when relative earnings are excluded from the equation.

We conclude that it is desirable to include the population in the receiving province as an explanatory factor in a projections equation. However, we feel somewhat uncertain about the population in the sending province. It may be worth excluding it from the equation altogether.

TABLE II.5: SELECTED MIGRATION EQUATIONS FOR 1983-84 - FEMALES AGED 18-24

FACTOR	EQUATION NUMBER											
	1	2	3	4	5	6	7	8	9	10	11	12
LMIGSRL	---	---	---	---	0.890*	0.893*	0.902*	0.891*	---	---	0.616*	0.599*
LPOPR	0.734*	0.645*	0.629*	0.563*	0.161*	0.132*	0.158*	0.146*	0.157*	0.032	0.139*	0.134*
LPOPS	0.567*	0.472*	0.487*	0.553*	-0.033	-0.007	-0.028	-0.018	0.004	0.146*	-0.042	-0.038
L(AVER/AVES)	-1.480	-1.235	-1.407	---	-0.418	---	-0.740	-0.581	-2.025*	---	-0.946*	-1.259*
UNR - UNS	-0.034	-0.097	---	-0.113	-0.092*	-0.098*	-0.007	---	-0.101*	-0.128*	-0.095*	-0.026*
DISTSR	-0.691*	-0.772*	-0.772*	-0.772*	-0.170*	-0.168*	-0.155*	-0.170*	-0.190*	-0.200*	-0.151*	-0.139*
EMPR/LFR	---	1.616	8.141*	-0.803	-5.579*	-6.416*	---	0.642	-2.205	-6.076	-4.705	---
EMPS/LFS	---	10.710	4.185	13.129	6.761*	7.559*	---	0.533	3.751	7.813*	5.538*	---
LMIGRSL	---	---	---	---	---	---	---	---	0.861*	0.846*	0.302*	0.329*
CONSTANT	0.435	-8.853	-8.853	-8.853	-0.834	-0.805	0.052	-0.829	-1.096	-1.233	-0.582	0.041
R^2	0.66	0.72	0.72	0.71	0.97	0.97	0.96	0.96	0.95	0.94	0.97	0.97

NOTES

An asterisk (*) denotes a statistically significant coefficient.

"L" before a factor denotes Log
"L" after a factor denotes the value in the previous year
MIGSR(MIGRS) denotes migration from sending (receiving) province to receiving (sending) province
POPR(POPS) denotes population in receiving (sending) province
AVER(AVES) denotes average earnings in receiving (sending) province
UNR(UNS) denotes unemployment rate in receiving (sending) province
DISTSR denotes distance between sending province and receiving province
EMPR(EMPS) denotes employment in receiving (sending) province
LFR(LFS) denotes Labour Force in receiving (sending) province

TABLE II.6: SELECTED MIGRATION EQUATIONS FOR 1983-84 - FEMALES AGED 25-44

FACTOR	EQUATION NUMBER											
	1	2	3	4	5	6	7	8	9	10	11	12
LMIGSRL	---	---	---	---	0.981*	0.985*	0.962*	0.974*	---	---	0.872*	0.728*
LPOPR	0.734*	0.668*	0.668*	0.570*	0.065*	0.042	0.089*	0.071*	0.160*	0.005	0.070	0.097*
LPOPS	0.565*	0.515*	0.515*	0.613*	-0.043	-0.024	-0.047	-0.039	-0.058	0.137*	-0.051	-0.063
L(AVER/AVES)	-1.525	-1.355	-1.402	---	-0.293	---	-0.594*	-0.370	-2.444*	---	-0.544	-1.084*
UNR - UNS	-0.009	-0.101	---	-0.108	-0.147*	-0.149*	-0.004	---	-0.070	-0.085	-0.138*	-0.002
DISTSR	-0.704*	-0.779*	-0.779*	-0.779*	-0.083*	-0.069*	-0.089*	-0.077*	-0.108*	-0.133*	-0.069*	-0.082*
EMPR/LFR	---	-0.732	11.674*	-4.456	20.222*	21.094*	---	-1.960	-6.537	12.923*	18.768*	---
EMPS/LFS	---	22.912	10.506*	26.636	18.935*	19.712*	---	0.852	6.521	13.748*	17.367*	---
LMIGRSL	---	---	---	---	---	---	---	---	0.029*	0.894*	0.114	0.243*
CONSTANT	-0.035	-19.215	-19.215	-19.215	1.227	1.311	0.112	1.071	0.036	-0.694	1.317	0.091
R^2	0.72	0.77	0.77	0.77	0.99	0.99	0.98	0.98	0.97	0.95	0.99	0.98

NOTES

An asterisk (*) denotes a statistically significant coefficient

"L"	before a factor denotes Log
"L"	after a factor denotes the value in the previous year
MIGSR(MIGRS)	denotes migration from sending (receiving) province to receiving (sending) province
POPR(POPS)	denotes population in receiving (sending) province
AVER(AVES)	denotes average earnings in receiving (sending) province
UNR(UNS)	denotes unemployment rate in receiving (sending) province
DISTSR	denotes distance between sending province and receiving province
EMPR(EMPS)	denotes employment in receiving (sending) province
LFR(LFS)	denotes Labour Force in receiving (sending) province

TABLE II.7: SELECTED MIGRATION EQUATIONS FOR 1983-84 - FEMALES AGED 45-64

FACTOR	EQUATION NUMBER											
	1	2	3	4	5	6	7	8	9	10	11	12
LMIGSRL	---	---	---	---	0.828*	0.830*	0.847*	0.826*	---	---	0.724*	0.722*
LPOPR	0.667*	0.607*	0.605*	0.575*	0.179*	0.141*	0.174*	0.172*	0.152	0.093	0.148*	0.140*
LPOPS	0.695*	0.637*	0.639*	0.668*	0.079	0.116*	0.076	0.089	0.248*	0.309*	0.078	0.077
L(AVER/AVES)	-0.463	-0.466	-0.401	---	-0.557	---	-0.410	-0.310	-0.879	---	-0.606	-0.501
UNR - UNS	-0.012	-0.018	---	0.002	-0.071	-0.046	-0.011	---	-0.014	9.925	-0.062	-0.015
DISTSR	-0.790*	-0.850*	-0.858*	-0.857*	-0.169*	-0.168*	-0.148*	-0.171*	-0.266*	-0.268*	-0.147*	-0.129*
EMPR/LFR	---	10.886	13.150*	12.601	-6.325	-4.297	---	2.429	5.437	8.696	-5.114	---
EMPS/LFS	---	12.757	10.494*	11.051	9.303	7.240	---	0.609	0.713	-2.492	7.390	---
LMIGRSL	---	---	---	---	---	---	---	---	0.705*	0.703*	0.131	0.151
CONSTANT	-1.226	22.445*	22.446*	22.452*	-3.054	-3.022	-0.368	-3.110	-6.010	-6.061	-2.399	-0.344
R^2	0.70	0.75	0.75	0.75	0.94	0.94	0.94	0.94	0.89	0.89	0.95	0.95

NOTES

An asterisk (*) denotes a statistically significant coefficient

"L" before a factor denotes Log
"L" after a factor denotes the value in the previous year
MIGSR(MIGRS) denotes migration from sending (receiving) province to receiving (sending) province
POPR(POPS) denotes population in receiving (sending) province
AVER(AVES) denotes average earnings in receiving (sending) province
UNR(UNS) denotes unemployment rate in receiving (sending) province
DISTSR denotes distance between sending province and receiving province
EMPR(EMPS) denotes employment in receiving (sending) province
LFR(LFS) denotes Labour Force in receiving (sending) province

TABLE II.8: SELECTED MIGRATION EQUATIONS FOR 1983-84 - MALES AGED 18-24

FACTOR	EQUATION NUMBER											
	1	2	3	4	5	6	7	8	9	10	11	12
LMIGSRL	---	---	---	---	0.903*	0.905*	0.897*	0.910*	---	---	0.637*	0.762*
LPOPR	0.678*	0.644*	0.634*	0.585*	0.117*	0.084*	0.114*	0.110*	0.042	-0.040	0.075	0.090*
LPOPS	0.696*	0.657*	0.667*	0.716*	0.038	0.068	0.038	0.036	0.145*	0.233*	0.052	0.043
L(AVER/AVES)	-0.542	-0.609	-0.443	---	-0.324	---	-0.306	-0.271	-0.886*	---	-0.498	-0.381
UNR - UNS	-0.016	-0.287	---	-0.269	-0.089	-0.079	-0.012	---	-0.480*	-0.452*	-0.211*	-0.014
DISTSR	-0.628*	-0.664*	-0.664*	-0.664*	-0.180*	-0.179*	-0.189*	-0.177*	-0.194*	-0.197*	-0.169*	-0.182*
EMPR/LFR	---	-11.859	3.469	-11.434	-4.587	-4.348	---	0.132	25,287*	24,596*	22.129*	---
EMPS/LFS	---	17.548	2.219	17.123	3.721	3.649	---	-1.044	24.610*	23,954*	10.107*	---
LMIGRSL	---	---	---	---	---	---	---	---	0.877*	0.872*	0.288*	0.148
CONSTANT	0.088	-3.951*	-3.951*	-3.951*	0.615	0.624	0.0	0.646	0.483	0.459	0.723	-0.001
R^2	0.70	0.72	0.72	0.72	0.97	0.97	0.97	0.97	0.96	0.95	0.98	0.97

NOTES

An asterisk (*) denotes a statistically significant coefficient

"L" before a factor denotes Log
"L" after a factor denotes the value in the previous year
MIGSR(MIGRS) denotes migration from sending (receiving) province to receiving (sending) province
POPR(POPS) denotes population in receiving (sending) province
AVER(AVES) denotes average earnings in receiving (sending) province
UNR(UNS) denotes unemployment rate in receiving (sending) province
DISTSR denotes distance between sending province and receiving province
EMPR(EMPS) denotes employment in receiving (sending) province
LFR(LFS) denotes Labour Force in receiving (sending) province

TABLE II.9: SELECTED MIGRATION EQUATIONS FOR 1983-84 - MALES AGED 25-44

FACTOR	EQUATION NUMBER											
	1	2	3	4	5	6	7	8	9	10	11	12
LMIGSRL	---	---	---	---	0.937*	0.947*	0.942*	0.939*	---	---	0.758*	0.826*
LPOPR	0.734*	0.723*	0.696*	0.578*	0.107*	0.040	0.096*	0.100*	0.169*	-0.033	0.107*	0.089*
LPOPS	0.587*	0.503*	0.530*	0.648*	-0.073*	-0.018	-0.067	-0.078	-0.091	9.156*	-0.088*	-0.071*
L(AVER/AVES)	-1.163	-1.497	-1.273	---	-0.631*	---	-0.553	-0.585*	-2.332*	---	-0.973*	-0.708*
UNR - UNS	-0.009	0.118	---	0.051	0.024	-0.005	-0.005	---	0.209*	0.102	0.061	-0.006
DISTSR	-0.681*	-0.743*	-0.743*	-0.743*	-0.127*	-0.120*	-0.123*	-0.126*	-0.149*	-0.173*	-0.119*	-0.118*
EMPR/LFR	---	26.129	11.173*	14.277	3.899	-1.254	---	0.899	29.581*	11.247	8.881	---
EMPS/LFS	---	-7.081	7.875*	4.771	-3.493	1.466	---	-0.526	28.490*	-9.438	-8.705	---
LMIGRSL	---	---	---	---	---	---	---	---	0.902*	0.866*	0.191*	0.124
CONSTANT	-0.084	16.710*	16.710*	16.710*	-0.015	0.159	0.342	0.015	-0.628	-1.271	0.191	0.346
R^2	0.71	0.76	0.76	0.75	0.98	0.98	0.98	0.98	0.96	0.94	0.98	0.98

NOTES

An asterisk (*) denotes a statistically significant coefficient

"L" before a factor denotes Log
"L" after a factor denotes the value in the previous year
MIGSR(MIGRS) denotes migration from sending (receiving) province to receiving (sending) province
POPR(POPS) denotes population in receiving (sending) province
AVER(AVES) denotes average earnings in receiving (sending) province
UNR(UNS) denotes unemployment rate in receiving (sending) province
DISTSR denotes distance between sending province and receiving province
EMPR(EMPS) denotes employment in receiving (sending) province
LFR(LFS) denotes Labour Force in receiving (sending) province

TABLE II.10: SELECTED MIGRATION EQUATIONS FOR 1983-84 - MALES AGED 45-64

FACTOR	EQUATION NUMBER											
	1	2	3	4	5	6	7	8	9	10	11	12
LMIGSRL	---	---	---	---	0.877*	0.869*	0.869*	0.877*	---	---	0.789*	0.797*
LPOPR	0.626*	0.590*	0.589*	0.564*	0.168*	0.108*	0.177*	0.167*	0.024	0.024	0.120	0.139*
LPOPS	0.698*	0.685*	0.685*	0.710*	0.043	0.110*	0.039	0.042	0.321*	0.321*	0.054	0.048
L(AVER/AVES)	-0.195	-0.279	-0.105	---	-0.692*	---	-0.651	-0.546	0.007	---	-0.585	-0.561
UNR - UNS	0.004	-0.057	---	-0.051	-0.048	-0.031	-0.006	---	-0.067	-0.067	-0.048	-0.002
DISTSR	-0.868*	-0.890*	-0.982*	-0.891*	-0.194*	-0.201*	-0.203*	-0.195*	-0.276*	-0.276*	-0.174*	-0.190*
EMPR/LFR	---	-0.358	6.749	-0.128	-6.744	-6.091	---	-0.862	-10.382	-10.386	-7.831	---
EMPS/LFS	---	16.971	9.781*	16.708	5.089	4.567	---	-0.846	11.136	11.142	5.410	---
LMIGRSL	---	---	---	---	---	---	---	---	0.770*	0.771*	0.124	0.097
CONSTANT	-0.767	15.952*	15,861*	15.916*	1.340	1.236	-0.167	1.404	-0.962	-0.963	2.052	-0.156
R^2	0.73	0.76	0.75	0.76	0.95	0.94	0.94	0.94	0.89	0.89	0.94	0.94

NOTES

An asterisk (*) denotes a statistically significant coefficient

"L"	before a factor denotes Log
"L"	after a factor denotes the value in the previous year
MIGSR(MIGRS)	denotes migration from sending (receiving) province to receiving (sending) province
POPR(POPS)	denotes population in receiving (sending) province
AVER(AVES)	denotes average earnings in receiving (sending) province
UNR(UNS)	denotes unemployment rate in receiving (sending) province
DISTSR	denotes distance between sending province and receiving province
EMPR(EMPS)	denotes employment in receiving (sending) province
LFR(LFS)	denotes Labour Force in receiving (sending) province

The estimated coefficients of unemployment rates, earnings and employment/labour force ratios in the equations which include lagged in-migrants and out-migrants provide some support for including these factors in the equation for making projections. But this support is far from overwhelming. The entries in Tables II.7 and II.10 suggest that employment and unemployment are not particularly relevant in projecting migration for those aged 45-64. Those in Tables II.5, II.6 and II.8 suggest that these factors have a more pronounced effect for those aged 18-24 and for females aged 25-44.

Some of our estimates suggest that it may not be particularly useful to include provincial differences in unemployment rates in an equation for making projections. Some experimentation should be undertaken to develop an indicator of unemployment that reflects the view of potential migrants in different provinces. Such an indicator should also probably be age-specific as well as sex-specific.

Similarly, some of our results do not support the use of relative average earnings as a factor in an equation for making projections of aggregate migration flows. In some equations, the coefficients for relative earnings are negative, suggesting that migration is higher to provinces with lower earnings. As noted in Section II.3, it may be important to take account of provincial differences in earnings for different education groups since migration varies considerably by educational level, but the data required to do so are not available at present.

c) Intertemporal stability

Our examination of the estimated coefficients in the regression equations for 1983-84, suggest that we should exclude some of our explanatory factors from our equation for making projections. However, since these results are based only on data for one year, it is useful to carry out a similar analysis using the estimated coefficients for other years to see if our conclusions hold. It is also important to determine whether or not the estimated coefficients change markedly from one year to the next, since the usefulness of a given equation for making projections may be affected by the stability of the coefficients over time.

The entries in Table II.11 to II.16 inclusive show our regression estimates for different years. Three model specifications are included in the tables which show separate equations for males and females in each of the three age-groups. The first specification (Table II.11 and II.12) excludes lagged out-migration and

in-migration and it is based on data from 1983-84, 1982-83, 1980-81 and 1978-79. The second specification (Tables II.13 and II.14) includes the lagged migration variables and all of the economic explanatory factors discussed above: the results are reported for 1983-84, 1982-83 and 1980-81. The final specification includes the lagged migration variables but it excludes employment/labour force ratios: the results are based on data for 1983-84, 1982-83 and 1980-81.

Before considering the stability of the coefficients over time, it is important to note that the goodness-of-fit is much the same for every year. Specifications without the lagged migration variables yield values of R-squared between 0.68 and 0.81, while the values of R-squared increase to 0.94 and over for the other specifications. In terms of aggregate measures, the goodness-of-fit is not noticeably better in one year than another.

It is clear from the entries in Tables II.11 and II.12 (which exclude the lagged migration variables) that the estimated coefficients for distance are reasonably stable except for those aged 45-64 in 1978-79. Thus, our results for 1983-84 are confirmed by the results for the other years, and we conclude there is a well-established pattern in the effects of distance, with relatively larger effects for those aged 45-64 and for females aged 18-24. The differences in the coefficients for the various age-groups are not pronounced, but they are large enough to justify the use of age-specific and sex-specific weights in deriving projections.

It is also clear that the results in these two tables confirm the importance of population in the sending and receiving provinces as standardization factors, although the coefficients for those aged 45-64 in 1978-79 are somewhat out of line with those for the other years.

The estimated coefficients for the economic factors for the other years support our conclusions using data for 1983-84, and the effects of these factors remain unclear. The coefficients for relative earnings are particularly interesting since the sign changes for most groups between 1980-81 and 1982-83. As discussed above, the instability in these coefficients may be due to the fact that it is not appropriate to use average earnings in each province to estimate the effects of earnings differences on decisions to migrate. As an alternative it may be useful to add another factor which is related to the level of economic activity in the country as a whole in each year, to see if the changes in the coefficients depend on general economic conditions.

When we examine the coefficients for the equations which include the lagged migration variables, (Tables II.13 to II.16), the importance of introducing a dynamic element in the equations becomes apparent. The coefficients for lagged out-migration and in-migration are always positive and statistically significant, and the coefficient for the former always exceeds that for the latter. However, there is some instability in the size of the estimated coefficients and this supports our view that there is a need for further exploratory analysis of the dynamic elements in migration flows. As discussed above, one possibility would be to experiment with the interaction of economic factors and lagged migration flows.

Our earlier conclusions concerning the uncertain impact of economic factors and their reformulation are still relevant when we examine the results for the equations which include lagged migration. As noted above, a scaling factor needs to be applied to the coefficients of other factors when comparisons are to be made among the coefficients of factors in equations with and without lagged migration. Comparison of the coefficients in Tables II.13 to II.16 is further complicated by the fact that two lagged variables are included in the equations, and it is therefore particularly difficult to find a suitable scaling factor. As an approximation, we can scale the coefficients by dividing by (1 - the sum of the coefficients for the lagged variables).

At first sight, the change of sign of some of the coefficients (between 1980-81 and 1982-83) when the lagged migration variables are included, suggests that they are highly unstable. However, we note that the sum of the coefficients is close to, or exceeds, 1 so that the change in sign may be eliminated when the scaling factor is applied. Thus, the changes in signs of the coefficients in Tables II.13 to II.16 do not necessarily indicate that the coefficients are unstable over time.

TABLE II.11: EQUATION EXCLUDING LAGGED FACTORS FOR FEMALE MIGRATION IN SELECTED YEARS

FACTOR	FEMALES AGED -24				FEMALES AGED 25-44				FEMALES AGED 45-64			
	83-84	82-83	80-81	78-79	83-84	82-83	80-81	78-79	83-84	82-83	80-81	78-79
LMIGSRL	---	---	---	---	---	---	---	---	---	---	---	---
LPOPR	.645*	.401*	.437*	.399*	.668*	.699*	.476*	.396*	.607*	.362*	.557*	.215
LPOPS	.472*	.452*	.584*	.772*	.515*	.605*	.705*	.775*	.637*	.562*	.642*	.961*
L(AVER/AVES)	-1.235	-1.067	.843	2.007	-1.355	-1.179	1.383	1.913	-.466	-.012	-.665	3.361*
UNR - UNS	0.097	.055	.049	-.004	.101	-.019	.007	.002	-.018	.049	-.156	.008
DISTSR	-.772*	-.734*	-.741*	-.781*	-.779*	-.639*	-.720*	-.767*	-.858*	-.858*	-.811*	-1.013*
EMPR/LFR	1.616	15.933	15.146	5.757	-.732	-.008	2.028*	10.181	10.886	27.116*	-2.743	8.236
EMPS/LFS	10.710	2.870	1.001	3.819	22.912	.008	1.311*	7.820	12.757	9.136	26.648	6.195
LMIGRSL	---	---	---	---	---	---	---	---	---	---	---	---
CONSTANT	-8.853	-13.650	-11.740	-6.582	19.215	0.097	-2.437	015.418	-22.445	-33.058	-22.651	-13.143
R^2	0.72	0.72	0.68	0.70	0.77	0.71	0.76	0.76	0.75	0.75	0.73	0.77

NOTES

An asterisk (*) denotes a statistically significant coefficient

"L" before a factor denotes Log
"L" after a factor denotes the value in the previous year
MIGSR(MIGRS) denotes migration from sending (receiving) province to receiving (sending) province
POPR(POPS) denotes population in receiving (sending) province
AVER(AVES) denotes average earnings in receiving (sending) province
UNR(UNS) denotes unemployment rate in receiving (sending) province
DISTSR denotes distance between sending province and receiving province
EMPR(EMPS) denotes employment in receiving (sending) province
LFR(LFS) denotes Labour Force in receiving (sending) province

TABLE II.12: EQUATION EXCLUDING LAGGED FACTORS FOR MALE MIGRATION IN SELECTED YEARS

FACTOR	MALES AGED -24				MALES AGED 25-44				MALES AGED 45-64			
	83-84	82-83	80-81	78-79	83-84	82-83	80-81	78-79	83-84	82-83	80-81	78-79
LMIGSRL	---	---	---	---	---	---	---	---	---	---	---	---
LPOPR	.644*	.484*	.401*	.461*	.723*	.522*	.719*	.444*	.590*	.400*	.236*	.262*
LPOPS	.657*	.611*	.790*	.833*	.503*	.561*	2.023*	.757*	.685*	.585*	.737*	.961*
L(AVER/AVES)	-.609	-.052	2.274*	1.553	-1.497	-.809	1.829*	2.448	-.279	.314	2.194*	2.887*
UNR - UNS	-.287	-.161	.055	-.035	.118	.114	.015	.020	-.057	.238	.058	.012
DISTSR	-.664*	-.600*	-.568*	-.648*	-.743*	-.700*	-.729*	-.741*	-.890*	-.846*	-.818*	-1.044*
EMPR/LFR	-11.859	- 3.025	8.092	2.138	26.129	27.184	-.247*	7.035	-.358	56.660*	25.748*	9.833
EMPS/LFS	17.548	14.388	1.957	4.253	-7.081	-1.802	-.227*	2.387	16.971	-16.285	9.575	10.354
LMIGRSL	---	---	---	---	---	---	---	---	---	---	---	---
CONSTANT	-3.951	-8.131	-7.327	-4.406	-16.710	-22.506	-.934	-7.679	-15.952	-37.556	-32.554	-18.923
R^2	0.72	0.72	0.68	0.70	0.76	0.77	0.87	0.74	0.76	0.78	0.77	0.81

NOTES

An asterisk (*) denotes a statistically significant coefficient

"L"	before a factor denotes Log
"L"	after a factor denotes the value in the previous year
MIGSR(MIGRS)	denotes migration from sending (receiving) province to receiving (sending) province
POPR(POPS)	denotes population in receiving (sending) province
AVER(AVES)	denotes average earnings in receiving (sending) province
UNR(UNS)	denotes unemployment rate in receiving (sending) province
DISTSR	denotes distance between sending province and receiving province
EMPR(EMPS)	denotes employment in receiving (sending) province
LFR(LFS)	denotes Labour Force in receiving (sending) province

TABLE II.13: EQUATION INCLUDING LAGGED FACTORS FOR FEMALE MIGRATION IN SELECTED YEARS

FACTOR	FEMALES AGED -24			FEMALES AGED 25-44			FEMALES AGED 45-64		
	83-84	82-83	80-81	83-84	82-83	80-81	83-84	82-83	80-81
LMIGSRL	0.616*	0.636*	1.007*	0.872*	0.561*	0.783*	0.724*	0.707*	0.821*
LPOPR	0.139*	0.046	0.028	0.070*	0.118*	-0.064	0.148*	0.075	-0.070
LPOPS	-0.042	0.010	-0.056	-0.051	0.003	0.001	0.078	0.063	0.164
L(AVER/AVES)	-0.946*	-1.385*	-0.415	-0.544	-1.191*	0.573	-0.606	-0.477	2.203
UNR - UNS	-0.095*	-.048	-0.043*	-0.138*	-0.011	0.005	-0.062	0.054	0.093
DISTSR	-0.151*	-0.035	0.079*	-0.068*	-0.090*	0.092*	-0.147*	-0.023	0.040
EMPR/LFR	-4.705	4.904	-2.318	18.768*	-0.013	0.268*	-5.114	6.554	14.717
EMPS/LFS	5.538*	-5.597	2.783	17.367*	-0.002	-0.167	7.390	-6.141	-7.416
LMIGRSL	0.302*	0.324*	0.034	0.114	0.369*	0.243*	0.131	0.193*	0.089
CONSTANT	-0.58	0.496	-0.614	1.317	-0.163	-0.021	-2.399	-0.677	-6.975
R^2	0.97	0.98	0.98	0.99	0.99	0.98	0.95	0.95	0.96

NOTES

An asterisk (*) denotes a statistically significant coefficient

"L"	before a factor denotes Log
"L"	after a factor denotes the value in the previous year
MIGSR(MIGRS)	denotes migration from sending (receiving) province to receiving (sending) province
POPR(POPS)	denotes population in receiving (sending) province
AVER(AVES)	denotes average earnings in receiving (sending) province
UNR(UNS)	denotes unemployment rate in receiving (sending) province
DISTSR	denotes distance between sending province and receiving province
EMPR(EMPS)	denotes employment in receiving (sending) province
LFR(LFS)	denotes Labour Force in receiving (sending) province

TABLE II.14: EQUATION INCLUDING LAGGED FACTORS FOR MALE MIGRATION IN SELECTED YEARS

FACTOR	MALES AGED -24			MALES AGED 25-44			MALES AGED 45-64		
	83-84	82-83	80-81	83-84	82-83	80-81	83-84	82-83	80-81
LMIGSRL	0.637*	0.612*	0.940*	0.758*	0.663*	0.797*	0.789*	0.449*	0.768*
LPOPR	0.075	0.048	-0.057*	0.107*	0.082*	-0.097	0.120	0.030	-0.077
LPOPS	0.052	0.100*	0.025	-0.088*	0.051	0.006	0.054	0.167*	0.037
L(AVER/AVES)	-0.498	-0.462	0.388	-0.973*	-1.131*	0.798*	-0.585	0.082	0.503
UNR - UNS	-0.211*	-0.153	-0.018	0.061	0.133*	0.008	-0.048	0.198*	0.029
DISTSR	-0.169*	-0.087*	0.100*	-0.119*	-0.069*	0.115*	-0.174*	-0.207*	0.132*
EMPR/LFR	11.129*	-8.460	-0.556	8.881	16.755*	-0.004	07.831	33.772*	9.352*
EMPS/LFS	10.107*	8.669	1.942	-8.705	15.380*	0.002	5.410	27.755*	-1.537
LMIGRSL	0.288*	0.289*	0.093*	0.191*	0.268*	0.248*	0.124	0.327*	0.200*
CONSTANT	0.723	-0.297	-1.391	0.191	-1.596	0.073	2.052	-5.402	-7.307
R^2	0.98	0.97	0.99	0.98	0.99	0.99	0.94	0.96	0.97

NOTES

An asterisk (*) denotes a statistically significant coefficient

"L"	before a factor denotes Log
"L"	after a factor denotes the value in the previous year
MIGSR(MIGRS)	denotes migration from sending (receiving) province to receiving (sending) province
POPR(POPS)	denotes population in receiving (sending) province
AVER(AVES)	denotes average earnings in receiving (sending) province
UNR(UNS)	denotes unemployment rate in receiving (sending) province
DISTSR	denotes distance between sending province and receiving province
EMPR(EMPS)	denotes employment in receiving (sending) province
LFR(LFS)	denotes Labour Force in receiving (sending) province

TABLE II.15: EQUATION EXCLUDING EMPLOYMENT/LABOUR FORCE RATIOS FOR FEMALE MIGRATION IN SELECTED YEARS

FACTOR	FEMALES AGED -24			FEMALES AGED 25-44			FEMALES AGED 45-64		
	83-84	82-83	80-81	83-84	82-83	80-81	83-84	82-83	80-81
LMIGSRL	0.599*	0.644*	0.987*	0.728*	0.585*	0.815*	0.722*	0.730*	0.804*
LPOPR	0.134*	0.075	0.035	0.097*	0.117*	-0.064	0.140*	0.101	0.007
LPOPS	-0.038	-0.023	-0.060	-0.063	0.007	-0.007	0.077	0.036	0.084
L(AVER/AVES)	-1.259*	-1.388*	-0.539	-1.084*	-1.109*	0.493	-0.501	-0.710	0.661
UNR - UNS	-0.026*	-0.023*	-0.015*	-0.002	-0.006	-0.004	-0.014	0.007	-0.011
DISTSR	-0.139*	0.046	0.086*	-0.082*	-0.093*	0.102*	-0.129*	-0.018	0.098
EMPR/LFR	---	---	---	---	---	---	---	---	---
EMPS/LFS	---	---	---	---	---	---	---	---	---
LMIGRSL	0.329*	0.307*	0.058	0.243*	0.344*	0.220*	0.151	0.173*	0.155*
CONSTANT	0.041	-0.019	-0.260	0.091	-0.193	0.056	-0.344	-0.307	-0.417
R^2	0.97	0.98	0.99	0.98	0.99	0.98	0.95	0.95	0.96

NOTES

An asterisk (*) denotes a statistically significant coefficient

"L" before a factor denotes Log
"L" after a factor denotes the value in the previous year
MIGSR(MIGRS) denotes migration from sending (receiving) province to receiving (sending) province
POPR(POPS) denotes population in receiving (sending) province
AVER(AVES) denotes average earnings in receiving (sending) province
UNR(UNS) denotes unemployment rate in receiving (sending) province
DISTSR denotes distance between sending province and receiving province
EMPR(EMPS) denotes employment in receiving (sending) province
LFR(LFS) denotes Labour Force in receiving (sending) province

TABLE II.16: EQUATION EXCLUDING EMPLOYMENT/LABOUR FORCE RATIOS FOR MALE MIGRATION IN SELECTED YEARS

FACTOR	MALES AGED -24			MALES AGED 25-44			MALES AGED 45-64		
	83-84	82-83	80-81	83-84	82-83	80-81	83-84	82-83	80-81
LMIGSRL	0.762*	0.618*	0.940*	0.826*	0.621*	0.803*	0.797*	0.609*	0.812*
LPOPR	0.090*	0.059	-0.056	0.089*	0.072	-0.103*	0.139*	0.068	-0.048
LPOPS	0.043	0.090	0.020	-0.071*	0.060	0.006	0.048	0.119*	0.025
L(AVER/AVES)	-0.381	-0.691	0.437	-0.708*	-0.723*	0.803*	-0.561	-0.287	0.482
UNR - UNS	-0.014	0.001	0.005	-0.006	0.005	0.007	-0.002	0.010	-0.003
DISTSR	-0.182*	-0.083*	0.122*	-0.118*	-0.056	0.117*	-0.190*	-0.161*	0.204*
EMPR/LFR	---	---	---	---	---	---	---	---	---
EMPS/LFS	---	---	---	---	---	---	---	---	---
LMIGRSL	0.148	0.287*	0.113*	0.124	0.322*	0.245*	0.097	0.210*	0.212*
CONSTANT	-0.001	-0.146	-0.332	0.346	-0.398	0.077	-0.156	0.115	-0.333
R^2	0.97	0.97	0.99	0.98	0.99	0.99	0.94	0.95	0.97

NOTES

An asterisk (*) denotes a statistically significant coefficient

"L"	before a factor denotes Log
"L"	after a factor denotes the value in the previous year
MIGSR(MIGRS)	denotes migration from sending (receiving) province to receiving (sending) province
POPR(POPS)	denotes population in receiving (sending) province
AVER(AVES)	denotes average earnings in receiving (sending) province
UNR(UNS)	denotes unemployment rate in receiving (sending) province
DISTSR	denotes distance between sending province and receiving province
EMPR(EMPS)	denotes employment in receiving (sending) province
LFR(LFS)	denotes Labour Force in receiving (sending) province

III OCCUPATIONAL DISTRIBUTION OF MIGRATION FLOWS

III.1 Introduction

The data available on the occupations of individuals are more limited than those for aggregate migration flows so that it is not possible at this time to build a regression model for projecting the occupations of migrants. Instead we suggest that projections of the occupations of migrants be derived:

a) by making assumptions about the distribution of migrants across occupations; and
b) by applying these distributions to the aggregate migration flows projected using one or more of the equations discussed in the previous chapter.

In order to develop a reasonable basis for making appropriate assumptions about the occupational distributions of migrants, we analyzed data from the 1981 Census and from the Current Population Profile (CPP) which was carried out as a supplement to the Labour Force Survey in August 1982. In carrying out this analysis, we tried to answer three questions:

1. Is the occupational distribution of migrants substantially different from that for the total population?

2. Are the occupational distributions for male and female migrants markedly different?

3. Is the occupational distribution of migrants relatively stable over time?

The results of this analysis are discussed in this chapter.

III.2 Results

The entries in Tables III.1 through III.3 show the data from the 1981 Census and the CPP which we used in our analysis. In Table III.1, we compare the occupational distributions of the total population from the Census and from the CPP for males and females. Note that the residual occupation major group (occupation not elsewhere specified) is excluded since in the CPP this category includes individuals whose occupation could not be determined from the available information.

On the whole, the occupational distributions from the Census and from the CPP are fairly similar, but there are a few relatively large differences. For example, the proportion of males in managerial occupations is considerably higher for the Census than for the CPP. For both males and females, the proportion in service occupations also differs substantially between the two sources. Some proportion of these differences is clearly due to sampling error, but it is likely that some differences are due to differences in the methodology or coverage of the two sources of data.

In Table III.2, we compare the occupational distributions of all persons who had changed province since 1976 using data from the 1981 Census and from the CPP. The distributions here are also fairly similar, although as before there are some large differences. The discrepancy is again large for managerial and administrative occupations and for females in service occupations, and there are also relatively large differences for farming and for fabricating occupations. But on the whole the differences between the distributions for migrants are much the same as for the total population so that it seems reasonable to assume that the occupational distribution of migrants from the CPP is consistent with that from the Census.

It is also interesting to compare the data in Table III.2 with those in Table III.1 to determine how the occupational distribution of migrants differs from that of the total population. The main differences between the two distributions are the following: the proportion of migrants is higher than the proportion in the total population for managerial, natural science and service occupations; the proportion of migrants is lower than that in the total population for farming, processing and fabricating occupations.

Finally, in Table III.3, we use data from the CPP to compare the occupational distribution for persons who migrated since September 1981 with that for those who migrated since September 1980. Note that similar data were not collected in the Census. The occupational distributions for the two time periods are remarkably similar for both males and females, especially in view of the smaller numbers (and hence larger sampling errors) on which these proportions are based.

As above it is interesting to compare the occupational distribution of recent migrants (Table III.3) with that for long-term migrants (Table III.2). Again the proportions in the different occupations are remarkably similar, but there are two striking exceptions. For managerial and administrative occupations, the proportion for recent migrants exceeds that for long-term migrants; the reverse is the case for sales occupations.

There are of course many possible explanations for this difference. One possibility is that the pattern of return migration is very different for migrants in managerial and administrative and in sales occupations: the former may be much less likely (while the latter may be much more likely) than other migrants to return to their home province, so that the proportions of migrants in these occupations would change over time.

TABLE III.1: OCCUPATIONAL DISTRIBUTIONS (%) OF TOTAL POPULATION FROM CENSUS AND FROM CPP

OCCUPATION	MALES		FEMALES	
	CENSUS	CPP	CENSUS	CPP
Managerial	11.5	8.6	5.5	5.0
Natural Science	5.0	4.8	1.2	1.0
Social Science	1.3	1.3	2.1	2.0
Religion	0.4	0.3	0.2	0.1
Teaching	3.0	2.5	6.3	5.4
Medicine	1.7	1.6	8.7	7.9
Artistic	1.5	1.4	1.4	1.3
Clerical	7.2	6.5	36.5	32.0
Sales	8.7	10.1	9.4	10.0
Service	9.6	11.7	15.8	21.0
Farming	5.4	6.5	2.2	3.0
Fishing	0.6	0.6	0.0	0.1
Forestry	1.1	1.2	0.1	0.1
Mining	1.0	1.0	0.0	0.0
Processing	5.4	5.2	2.2	2.4
Machining	4.2	3.7	0.4	0.3
Fabricating	10.4	11.1	4.9	5.0
Construction	11.0	10.6	0.3	0.2
Transportation	6.3	5.9	0.6	0.5
Materials Handling	2.8	3.7	1.2	1.5
Other Crafts	1.7	1.6	0.6	0.6

Source: Census data obtained from the Statistics Canada Catalogue 92-917. CPP data obtained from special tabulations provided by the Special Surveys Division, Statistics Canada.

TABLE III.2: OCCUPATIONAL DISTRIBUTIONS (%) OF ALL MIGRANTS FROM CENSUS AND FROM CPP

OCCUPATION	MALES		FEMALES	
	CENSUS	CPP	CENSUS	CPP
Managerial	14.1	10.6	7.0	5.7
Natural Science	8.0	7.7	2.2	1.3
Social Science	1.7	1.3	2.8	2.5
Religion	0.8	0.3	0.4	0.2
Teaching	2.3	2.5	5.4	6.5
Medicine	1.8	1.6	10.0	7.7
Artistic	2.0	1.4	2.2	2.4
Clerical	6.0	5.7	37.4	29.8
Sales	8.7	9.4	8.9	10.8
Service	14.1	13.2	16.7	22.8
Farming	2.2	4.3	1.1	1.9
Fishing	0.3	0.3	0.0	0.1
Forestry	0.7	0.9	0.1	0.0
Mining	2.0	1.0	0.0	0.1
Processing	3.2	4.7	1.1	1.8
Machining	3.1	3.7	0.3	0.3
Fabricating	7.9	11.1	1.9	3.8
Construction	11.7	10.9	0.4	0.3
Transportation	5.5	4.9	0.7	0.5
Materials Handling	2.3	2.4	0.7	0.9
Other Crafts	1.2	1.2	0.7	0.5

Source: Census data obtained from Statistics Canada Catalogue 92-927. CPP data obtained from special tabulations provided by the Special Surveys Division of Statistics Canada.

TABLE III.3: OCCUPATIONAL DISTRIBUTIONS (%) OF MIGRANTS SINCE 1981 AND SINCE 1980 FROM CPP

OCCUPATION	MALES		FEMALES	
	1981	1980	1981	1980
Managerial	4.9	6.0	3.9	4.4
Natural Science	7.3	9.3	2.6	2.4
Social Science	1.7	1.5	1.5	1.7
Religion	0.4	0.3	0.0	1.0
Teaching	2.9	2.3	5.0	4.1
Medicine	1.9	2.7	7.2	8.0
Artistic	1.6	1.7	1.7	1.5
Clerical	5.2	4.8	36.5	35.9
Sales	12.1	11.0	12.4	11.0
Service	14.8	15.3	20.1	21.9
Farming	6.2	4.5	2.8	1.9
Fishing	0.5	0.5	0.0	0.0
Forestry	0.5	1.0	0.0	0.0
Mining	2.2	2.0	0.2	0.3
Processing	5.5	4.7	1.7	1.6
Machining	4.3	3.8	0.0	0.0
Fabricating	6.7	8.3	2.3	2.8
Construction	11.7	11.3	0.8	0.5
Transportation	6.9	5.6	0.2	0.1
Materials Handling	2.4	2.8	1.0	0.6
Other Crafts	0.3	0.3	0.0	0.0

Source: Data obtained from special tabulations from the CPP file, provided by Special Surveys Division, Statistics Canada.

IV CONCLUSIONS AND RECOMMENDATIONS

IV.1 Introduction

We had two objectives in carrying out this study: one was to develop exploratory models for making projections of interprovincial migration flows in aggregate, while the other was to develop a method for disaggregating these aggregate flows by occupation.

Our models for aggregate migration flows were developed using the familiar least-squares regression framework in which migration is expressed as a function of a number of explanatory factors. Since our models are based on aggregate data, we could not take account of the results of microeconomic studies on migration, which suggest that differences in the characteristics and experiences of individuals have a marked effect on migration decisions. Instead we linked migration flows to three main factors: economic conditions, geographic distance and population size. Annual data on interprovincial migration flows are available for male and female migrants in three age-groups, so that some demographic variables could be incorporated in our models. However, data are not available on other important characteristics (in particular on the educational qualifications and earnings of migrants), and these omissions limit the usefulness of our models.

Our models for projecting aggregate migration are based on the gravity model which has been widely used in research studies on regional migration. The gravity model assumes that migration flows are related to population size and to distance between provinces. We modified this basic model to include "economic distance", measured as the difference between unemployment rates and earnings in two provinces, as another independent variable.

There are a number of limitations in the models we developed for aggregate migration flows, and two in particular should be borne in mind in using our results. In the first place, our objective was to develop models for making projections of migration flows and not to provide the best estimates of the effects of the factors affecting such flows. If the latter had been our objective, then it would probably have been better to use data on the migration experience of individuals for developing the models. Aggregate data are more suitable for developing the projections models required for C.O.P.S.

In the second place, the aggregate data that were readily available to us, covered a relatively short time-period. Since we had data for only eight years, we could not explore the effects of a number of explanatory factors if we were to build a time-series model. We

therefore estimated our equations using cross-sectional data for each year, but this approach is not completely satisfactory since no direct account can be taken of time-series effects. In recent years some analysts have tried to improve on such cross-section estimates by pooling cross-sectional and time-series data so that the estimated coefficients will then include some time-series effects. However, the pooling procedure may yield worse results if the cross-section estimates vary substantially from one year to the next and if the sample is relatively small. Since this is the case for the models we developed, we decided not to use this procedure. We discuss this issue and some other technical considerations for further analysis in the Appendix: some relevant technical references are also provided here.

Data on the occupations of migrants are not available on an annual basis, so that our model for making projections by occupation is based on assumptions about the occupational distribution of migrants. These assumptions are based on data from the 1981 Census and from the Supplement to the Labour Force Survey carried out in August 1982.

Our conclusions have already been presented in the different sections of the report. However, we have brought together our more important conclusions and they are presented in the list below. The list also includes some suggestions for extending the analysis we have carried out.

IV.2 Aggregate migration flows

A. Method

(i) The linear regression model we have used, provides a suitably flexible approach for choosing the explanatory factors to be included in a projections model. However, some care is required in interpreting the estimated coefficients, especially when the factors are mutually interrelated and when the factors representing return migration and the stability of migration flows are included.

(ii) Cross-sectional data provide an adequate number of observations in any given year for developing experimental models. The range in the observations is relatively large, so that the influence of different explanatory factors can be explored. The exclusion of very small migration flows for some age-sex groups makes little difference to the results.

(iii) There may be some value in carrying out experiments in pooling the data for several years for estimation purposes. However some care is necessary because of a change in the age-groups covered in the migration statistics from 1981-82. The software for dealing with such pooled data is readily available.

(iv) The explanatory factors we used, can be expanded to include the interaction between different factors or non-linearities in their effects. This may help to improve the evident instability of the estimated coefficients.

B. Distance

(v) Our simple specification for distance is significant in almost all of our equations, and the signs of the coefficients are as expected. The size of the coefficient seems to increase with age, and the coefficients appear to be different for young males and for young females.

(vi) Our treatment of distance is symmetric in the sense that its recorded value is the same for, say, migration from Newfoundland to Alberta as for migration from Alberta to Newfoundland. Thus the direction of migration is assumed to be irrelevant. It would be interesting to explore the effects of asymmetries in the distance factor to try to pick up the effects of factors such as pre-retirement migration.

(vii) Our specification for distance is arbitrary, and it would be useful to explore the sensitivity of the estimated coefficients to changes in these assumptions.

C. Standardization for population size

(ix) The use of the sum of the population in the sending and in the receiving provinces for standardization, does not appear to be as satisfactory as the inclusion of the population in each province as a separate explanatory factor in the regression equation. The estimated coefficients of the two population factors differ from unity so that their values cannot really be specified beforehand. They also differ from each other and in many cases they differ across age-groups as well. The use of the employed population instead of the total population for standardization, has almost no effect on the results.

(x) At least one of the standardization factors is generally significant even when lagged migration factors are included. Our results indicate that it is important to include the population in the receiving province as an explanatory factor, but the case for including the population in the sending province is not as clear.

(xi) The estimated coefficients for the population in the receiving province are sensitive to the inclusion of relative earnings in the equations.

(xii) The interaction of population and economic factors needs further investigation. It also seems important that standardization for population differences when lagged migration factors are included, should be examined in more detail.

D. Lagged migration flows as explanatory factors

(xiii) The inclusion of lagged migration (both into and out of a given province) as factors in our equations can be interpreted in a meaningful way. One factor represents the stability of migration flows, and it permits a distinction between the long-run and short-run effects of economic factors on migration. The other represents an approximation for return migration. Both factors have a pronounced impact on the estimated coefficients in our regression models.

(xiv) The estimated coefficients show that neither of the two lagged migration factors is neutral with respect to the estimated coefficients of the economic factors. This is consistent with the view that economic factors have long-term effects which are perhaps obscured in those regression equations in which the lagged migration factors are omitted.

(xv) These long-term effects need more complete analysis than has been possible in this study. The additional analysis could include other exploratory regression models which have not been considered here. However, as noted below, the treatment of the economic factors may have to be modified first.

(xvi) The coefficients of lagged in-migration when lagged out-migration is excluded, are unreasonably high and this suggests that it makes more sense to include both lagged migration factors in the projections equation.

(xvii) The available evidence on return migration shows that although most return migration takes place in the first year after migration, some such migration occurs for several years afterwards. It would be useful to examine more recent data on return migration flows than were available to us, and to try to include such flows in a model.

E. "Economic distance"

(xviii) In the gravity model, economic factors in two regions are treated equally and the simple difference between the values of a factor is often used as a measure of the "economic distance" between the two regions. Our results suggest that this treatment may not be appropriate in the context of interprovincial migration. There is no support for the view that the "push" and "pull" of economic factors have an equal impact on migration flows.

(xix) Our empirical results are more consistent with a dynamic perspective that stresses the impact of the changing pulse of the overall level of economic activity. The common treatment of economic factors in two provinces is liable to generate "false" signs for the estimated coefficients because of these constraints. The overall effect may overwhelm the impact of interprovincial differentials and this suggests that we reject the familiar form of the gravity model.

(xx) The rejection of the use of simple differences between economic factors, for both unemployment and earnings, means that the specifications for the economic factors must be radically modified. This implies that a wider class of potential regression models should be explored.

(xxi) Economic conditions can nevertheless still be measured in terms of unemployment rates, earnings and the employment/labour force ratio. But the use of provincial averages (even for age-sex specific groups), may not provide a good approximation for the information that is of relevance to potential migrants. The modification of the use of averages should be explored by comparing data on the characteristics of migrants with data on the characteristics of non-migrants.

F. Personal characteristics

(xxii) There is enough heterogeneity in our estimated models to support the separation of the data into six age-sex categories. This can be interpreted as strong confirmation that personal characteristics are important in a projections moel. It suggests, moreover, that some attempt should be made to examine the effects of other personal characteristics, such as the educational qualifications of migrants. However, the data required to do so are not currently available.

(xxiii) Pooling of the cross-sectional data across age-sex groups within a given year can also be used to check the significance of differences among these groups. As noted above, adequate software is available to permit experimentation with the pooled data.

G. Temporal stability

(xxiv) Most of the models that we considered gave high values for conventional measures for goodness-of-fit. Unfortunately this is not sufficient to guarantee the stability of the estimated coefficients in the models, and there is clear evidence that the estimates are not stable over time.

(xxv) Since the temporal stability of the coefficients seems crucial for projections of future or hypothetical migration flows, there is a pressing need to see if parametric instability translates into a wide range of projections. Thus it would be useful to derive some projections by simulation, either within or outside the time-span of our data, to see how they compare with actual migration flows. Our basic recursive format is convenient for such simulation experiments, but the estimated coefficients should be "tuned" or "shocked" to assess sensitivity.

(xxvi) The temporal instability of the estimates may be due to the inherent complexity of real developments or to the use of annual aggregate data for understanding individual actions and decisions that are seldom made on an annual basis. However, the instability may also be due to inadequate specification of any of the influences discussed above. Modifications to standardization, to the treatment of economic factors, to distance measures or to the dynamics of migration, may all be a means of reducing the temporal instability of the coefficients. Experiments with these modifications, and having this instability in mind, are worthwhile. Such experimentation is relatively simple to pursue although the number of modifications considered at each stage must be limited.

IV.3 Occupational distribution of migrants

(xxvii) Our comparison of the occupational distributions of migrants using data from the 1981 Census and from the Labour Force Survey Supplement of August 1982, suggests that the two data sets are reasonably consistent with each other. Moreover, the proportion of migrants in each occupation appears to remain fairly stable over time, except for managerial and administrative occupations and for sales occupations. This suggests that the use of the occupational distribution of migrants in the 1981 census may provide a reasonable approximation for projection purposes.

Because of the differences in the proportion of migrants in managerial and administrative occupations and in sales occupations for different periods of migration, it would be worthwhile carrying out some experiments to determine the sensitivity of the occupational projections to changes in these proportions.

It may well be the case that relatively large changes in the proportions have only a small effect on the final occupational projections in any province.

IV.4 Final comments

(xxix) Our overall conclusion is that our findings provide a reasonable basis for developing projections of inter-provincial migration flows. Our findings provide strong support for our basic methodology, despite the difficulties we have noted and the additional analysis we have suggested. The regression framework provides a short list of explanatory factors which, after further experimentation, may provide a satisfactory basis for making projections.

REFERENCES

Ahamad, B. (1984):

A METHODOLOGY FOR PROJECTING INTERPROVINCIAL MIGRATION FLOWS BY OCCUPATION, Report prepared for Employment and Immigration Canada, (mimeo.)

Grant, E. Kenneth and John Vanderkamp, (1976):

THE ECONOMIC CAUSES AND EFFECTS OF MIGRATION: CANADA, 1965-71, Ottawa, Economic Council of Canada

Green, C, (1977):

"THE EMPLOYMENT RATIO AS AN INDICATOR OF AGGREGATE DEMAND PRESSURE", Monthly Labor Review, pp. 25-32

Greenwood, Michael J. (1975):

RESEARCH ON INTERNAL MIGRATION IN THE UNITED STATES: A SURVEY", Journal of Economic Literature, XIII,2

Moore, G. (1973):

HOW FULL IS FULL EMPLOYMENT?, Washington, American Enterprise Institute

Norris, Douglas (1983):

"NEW SOURCES OF CANADIAN SMALL AREA MIGRATION DATA", Review of Public Data Use, 11, January

Norris, Douglas and Linda Standish (1983):

A TECHNICAL REPORT ON THE DEVELOPMENT OF MIGRATION DATA FROM ADMINISTRATIVE RECORDS, Ottawa, Statistics Canada

Statistics Canada (1983):

THE LABOUR FORCE, May 1983, (Catalogue 71-001), Ottawa

Statistics Canada (1984):

EARNINGS OF MEN AND WOMEN 1981 AND 1982, (Catalogue 13-577), Ottawa

Statistics Canada (1985):

POPULATION PROJECTIONS FOR CANADA, PROVINCES AND TERRITORIES 1982-2006, (Cataogue 91-520), Ottawa

Vanderkamp, John (1973):

MOBILITY BEHAVIOUR IN THE CANADIAN LABOUR FORCE, Ottawa, Economic Council of Canada

APPENDIX: SOME TECHNICAL CONSIDERATIONS FOR FURTHER ANALYSIS

Introduction

The probability assumptions underlying the regression models we fitted for aggregate migration flows have not been discussed in the text. We therefore provide a brief discussion of the framework we used and some technical considerations for further analysis in this appendix.

In carrying out this study, our approach was to develop exploratory models for making projections of aggregate migration flows rather than to confirm an established structure in the determination of these flows. The familiar elements of the classical linear model provide the best assumptions for developing such exploratory models. Although these assumptions have to be modified when the dynamic elements (in the form of lagged migration) are included in our models, the classical linear model provides an "ideal" framework if test statistics are to be given their conventional interpretation. The consequences of the probability assumptions and the feasibility of developing more complex statistical models or estimation procedures are discussed in this appendix.

The Ideal Framework

Suppose that $y(g,t)$ contains 90 basic observations on interprovincial migration for a given age-sex group ($g=1,2,..,G$) of the Canadian population in a given year ($t=1,2,...,T$). The corresponding explanatory factors, or "carriers", are recorded in a matrix $x(g,t)$ and the errors in the equation are contained in $u(g,t)$. Then the standard linear framework is given by:

$$y(g,t) = x(g,t)b(g,t) + u(g,t)$$

where $b(g,t)$ are unknown weights for the influence of the explanatory factors. These weights, or parameters, are (as we noted when we discussed our estimates in the text) potentially different across the demographic groups. They may also be different, or unstable, across years because the linear approximation we use is too restrictive, or because the explanatory factors included are not completely appropriate.

This potential instability does not matter within a basic file of 90 observations, since it does not affect the homogeneity of the 90 observations. However, it does matter if basic files are pooled across the age-sex groups or over time. For the basic files (of unpooled data) we assume that the weights associated with particular factors are unaffected by the choice of origin and destination of migration. These weights are assumed to be "fixed" and the observations for the factors are assumed to be either non-stochastic or to be uncorrelated with the errors in the equation. When the two dynamic terms (lagged migration) are included in the equation, this assumption must be amended so that the observations are not correlated with the coterminous errors

only. In both of the above cases, least-squares estimates are unbiased or asymptotically unbiased provided that the error terms have zero means.

For carrying out statistical tests, it must also be assumed that the errors are homoscedastic (that is, that they have constant variance across the group) and are free from autocorrelation. In this case, the use of the familiar confidence intervals, estimated standard errors and F-statistics is valid provided that the errors are normally distributed.

The validity of these assumptions can be explored using diagnostic procedures that have been developed in recent years. The obvious steps may include looking for patterns in the residuals obtained from the least-squares fits. Such patterns may indicate that the assumptions about homoscedasticity and autocorrelation are violated. They may also permit us to identify observations that have a particularly strong influence, and these may be related to specific factors in some provinces which are not explicitly recognized in our models.

Adjustments of separate basic files

When inspection of residuals or other prior considerations suggest that the "ideal" considerations are invalid, some adjustments or "corrections" may be considered. For example, we may use a transformation to correct for heteroscedasticity or spatial autocorrelation. The essential element in these transformations is that functions of the least-squares residuals are used to scale the observations, or to form differences of them, before re-estimating the weights for the explanatory factors. Unfortunately such corrections adversely affect the probability properties of the observations for the explanatory factors so that they may introduce biases. In large samples, there is reason to hope that these biases will be small, but it is not clear that 90 observations are large enough for this to be the case. There is nevertheless a case to be made for some experimentation of this type.

Pooling of the data

The equations derived for the basic files can be based on pooled data in three ways. They can be linked (i) across groups for a given year, or (ii) across years for a given group, or (iii) across both groups and years. Such pooling is limited only by computational difficulties and by the need to specify the joint distribution of the pooled errors {u(g,t) for all g,t}. The statistical framework to be used with pooled data is a variant of the three-stage least-squares method developed by Zellner and Theil two decades ago. The SUR variant (or Seemingly Uncorrelated Regression model) can be fitted using SAS and other popular software packages. It requires the prior estimation of the dispersion matrix of the pooled errors (for given g or for given t). For our models, the chances of getting a good fit are higher for pooling across groups than across years, since there are simply

far too few years of data for us to obtain good estimates of the dispersion matrix of the errors. It is for this reason that our recommendations in the text are restricted to pooling across groups rather than over years; however, we do recommend that there should be further appraisal of the temporal instability of the estimated weights for the explanatory factors using separate basic files.

The use of the SUR procedure with small samples combined with constraints across equations, has recently been the subject of some criticism in the literature (see Fiebig and Theil, 1983, and Theil and Rosalsky, 1984). However, the impact of such criticism is reduced when the sample becomes large. Descriptions of the three-stage least-squares and SUR frameworks are provided in Chow (1983, Section 5.4) and Srivastava and Dwivedi (1979).

REFERENCES FOR APPENDIX

Chow, G. C. (1983):
ECONOMETRICS, New York, McGraw-Hill.

Fiebig, D. G. and H.Theil (1983):
"The Two-Perils of Symmetry-Constrained Estimation of Demand Systems", Economics Letters, Vol. 13, pp. 105- .

Srivastava, V. K. and T. D. Dwivedi (1979):
"Estimation of Seemingly Unrelated Regression Equations: A Brief Survey", Journal of Econometrics, Vol. 10, pp. 15-32.

Theil, H. and M. C. Rosalsky (1984):
"More on Symmetry-Constrained Estimation", Economics Letters, Vol. 17, pp. 257-263.